Digital neuropsychological assessment: Feasibility and applicability in patients with acquired brain injury

Lauriane A. Spreij , Isabel K. Gosselt , Johanna M. A. Visser-Meily & Tanja C. W. Nijboer

Routledge
Taylor & Francis Group

Digital neuropsychological assessment: Feasibility and applicability in patients with acquired brain injury

Lauriane A. Spreij [ID][a*], Isabel K. Gosselt [ID][a*], Johanna M. A. Visser-Meily [ID][a,b] and Tanja C. W. Nijboer [ID][a,c]

[a]Center of Excellence for Rehabilitation Medicine, UMC Utrecht Brain Center, University Medical Center Utrecht, and de Hoogstraat Rehabilitation, Utrecht, The Netherlands; [b]Department of Rehabilitation, Physical Therapy Science & Sports, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht, The Netherlands; [c]Department of Experimental Psychology, Helmholtz Institute, Utrecht University, Utrecht, The Netherlands

ABSTRACT

Introduction: Digital neuropsychological assessment (d-NPA) has several advantages over paper-and-pencil tests in neuropsychological assessment, such as a more standardized stimulus presentation and response acquisition. We investigated (1) the feasibility and user-experience of a d-NPA in patients with acquired brain injury (ABI) and healthy controls; (2) the applicability of conventional paper-and-pencil norms on digital tests; and (3) whether familiarity with a tablet would affect test performance on a tablet.

Method: We administered a d-NPA in stroke patients ($n = 59$), traumatic brain injury patients ($n = 61$) and healthy controls ($n = 159$). The neuropsychological tests were presented on a tablet and participants used a pencil stylus to respond. We examined the completion rate to assess the feasibility, and a semi-structured interview was conducted to examine the user-experience. The applicability of conventional norms was examined by the number of healthy controls performing <10th percentile, which was expected to be no more than 10%. The effect of tablet familiarity on test performance was examined with a regression-based model.

Results: Overall, 94% of patients completed the d-NPA. The d-NPA was considered pleasant by patients and healthy controls. Conventional norms that exist for paper-and-pencil tests were not applicable on the digital version of the tests, as up to 34% of healthy controls showed an abnormal performance on half of the tests. Tablet familiarity did not affect test performance on a tablet, indicating that participants who were more experienced with working with a tablet did not perform better on digital tests.

Conclusions: The administration of a d-NPA is feasible in patients with ABI. Familiarity with a tablet did not impact test performance, which is particularly important in neuropsychological assessment. Future research should focus on developing norms in order to implement a d-NPA in clinical practice.

Introduction

Neuropsychological paper-and-pencil tests are widely used to assess cognitive functioning. Their validity and reliability have been evaluated and documented thoroughly (International Test Commission, 2001; Lezak et al., 2004; Muñiz & Bartram, 2007). Over the last decades, computerized tests and test batteries have been developed to administer, score, and interpret measures of cognitive functioning (Kane & Kay, 1992; Parsey & Schmitter-Edgecombe, 2013; Rabin et al., 2014). Computerized tests have several advantages over paper-and-pencil tests, as they allow a more standardized stimulus presentation and response acquisition, automated scoring (which is cost and time efficient and less prone to errors), and a convenient data storage (Bauer et al., 2012; Cernich et al., 2007). Some

computerized test batteries translated conventional paper-and-pencil tests into computerized tests, and other test batteries developed new tests (see Supplementary Table 1 for an overview of computerized test batteries).

There are, however, several aspects that compromise the usability of computerized test batteries in clinical practice (Bauer et al., 2012; Bilder & Reise, 2019; Schlegel & Gilliland, 2007). For instance, introducing new tests in clinical practice requires clinicians to invest time in learning the structures, instructions and underlying constructs of the tests. In addition, norm scores of computerized test batteries are often not available (Canini et al., 2014; Schlegel & Gilliland, 2007). Furthermore, some test batteries allow self-

administration with minimal interaction between the clinician and patient. Important behavioral observations, such as fatigue or unexpected distractors, are therefore lost (Bilder & Reise, 2019; Harvey, 2012; Kaplan, 1988; Witt et al., 2013). Finally, an individual's familiarity with a response device (e.g., keyboard, computer mouse, joystick or touch-screen devices) may affect test performance (Germine et al., 2019). For instance, people with greater computer experience tend to perform better on computerized tests than those with less computer experience (Iverson et al., 2009; Tun & Lachman, 2010). Previous studies – where several response devices were compared – concluded that touch-screen devices are considered favorable in cognitive assessment, due to an intuitive and natural interaction (Canini et al., 2014; Carr et al., 1986; Findlater et al., 2013; Murata & Iwase, 2005). Since touch-screen devices require little training, little cognitive demands, and little hand-eye coordination, they have been considered especially suitable among people who are less exposed to technology (Canini et al., 2014; Cernich et al., 2007; Joddrell & Astell, 2016). However, further research is needed regarding the potential effect of familiarity with touchscreen devices on test performance (Germine et al., 2019; Jenkins et al., 2016; Joddrell & Astell, 2016; Wallace et al., 2019).

In this study, we investigated a digital neuropsychological assessment (d-NPA) containing twelve conventional paper-and-pencil tests that were translated to digital tests. The d-NPA was administered by a neuropsychologist so no behavioural observations would be lost. The digital tests were presented on a touch-screen device (i.e., tablet) and participants used a pencil stylus to respond. Our first aim was to investigate the feasibility and user-experience in patients with acquired brain injury (ABI) and healthy controls. This is important, as patients with ABI may experience sensory overload when using technological devices, in particular in demanding or stressful situations (Scheydt et al., 2017). In order to gain diagnosis-specific insights, we recruited patients with stroke and patients with traumatic brain injury (TBI), which are the most common causes of ABI. Second, as a paper-and-pencil administration differs from a digital administration, norms that exists for paper-and-pencil tests may not simply be applicable to digital versions of the tests, even though the structure, instructions and underlying constructs remain similar (Bauer et al., 2012; Germine et al., 2019; Parsey & Schmitter-Edgecombe, 2013). Therefore, we investigated the applicability of conventional norms that exist for paper-and-pencil tests on our digital versions of the tests. Conventional norms correct for an effect of age,

sex and/or level of education (Heaton & Matthews, 1986). However, technology-specific factors might impact test performance as well (American Psychological Association, 1986). Since familiarity with a particular response device seems to be an important factor (Germine et al., 2019; Jenkins et al., 2016), our third aim was to investigate whether familiarity with a tablet influenced test performance on a d-NPA and should be taken into account in future norms.

Methods

Participants

We recruited participants between April 2017 and April 2018. Stroke and TBI patients who were treated at the *University Medical Center Utrecht* or *De Hoogstraat Rehabilitation Center*, the Netherlands, between January 2015 and April 2018, were considered for inclusion. The inclusion criteria were: (1) clinically diagnosed stroke as indicated by clinical computed tomography (CT) or magnetic resonance imaging (MRI) scan, and clinically diagnosed TBI as indicated by a neurologist; (2) aged ≥18 years; (3) fluent in Dutch; and (4) living at home at the time of participation. We excluded patients with severe communication and/or language deficits (evaluated by researcher) to prevent unreliable test performances, as language deficits would hamper the understanding of test instructions and providing verbal responses. Eligible patients were invited to participate via an information brochure that was handed out by a clinician (e.g., rehabilitation specialist, occupational therapist) or send by post. The research session took place at the medical center, the rehabilitation center, or at a patient's home.

As a reference group, healthy controls were recruited among acquaintances of the researchers, via (sport) clubs, and via social media. The data of an additional group of healthy controls was obtained from Philips Research who conducted a similar study to enlarge the sample and its generalizability. These participants were recruited from a proprietary database of elderly people. Overall, the inclusion criteria of the healthy controls were: (1) aged ≥18 years; and (2) fluent in Dutch. We excluded participants with a medical history of neurological and/or psychiatric disorders for which medical treatment was necessary (based on self-report). All participants gave written informed consent. The research protocol of the current study was approved by the Medical Ethical Committee of the University Medical Center Utrecht (METC protocol number 16–760/C). The

study was performed in accordance with the Declaration of Helsinki.

Digital neuropsychological assessment (D-NPA)

A trained neuropsychologist (one licensed and four residents) administered the twelve tests of the d-NPA in a fixed order: Rey Auditory Verbal Learning Test (RAVLT) immediate recall, Trail Making Test (TMT) part A and B, Cube Drawing, O-Cancellation, Clock Drawing, Star Cancellation, RAVLT delayed recall and recognition, Rey-Osterrieth Complex Figure (ROCF) copy, Verbal Fluency Letter, ROCF immediate recall, Digit Span forwards and backwards, Verbal Fluency Category, Stroop Color and Word Test (Stroop), ROCF delayed recall, and the Wisconsin Card Sorting Test (WCST). See Supplementary Table 2 for references to the used stimuli, instructions and scoring, the outcome measures, and the conventional norms.

The software of the d-NPA was a research prototype developed by Philips Research (Vermeent et al., 2020). The software included test descriptions, test instructions, administration forms to record observations, and stimuli (auditory and visual). It was designed to be used on a regular laptop (HP© EliteBook840) in combination with a tablet (Apple© iPad Pro) with a screen size of 12.9-inch and a screen resolution of 2732 × 2048 pixels. Participants used a pencil stylus (Apple© Pencil) on the tablet to conduct drawing tests or tests that needed a manual response. A tablet was placed in front of the participant and the neuropsychologist sat across them while controlling the tests on a regular laptop. The brightness of the tablet screen and the volume of the laptop were set to 100%.

Verbal responses (RAVLT, Verbal Fluency, Digital Span, Stroop) were recorded by the audio recorder on the tablet and scored on the laptop during and/or after the administration by the neuropsychologist. Manual responses (O-cancellation, Star Cancellation, TMT, WCST) were recorded and scored automatically, but corrected based on observations of the neuropsychologist if necessary (e.g., if a non-target was unintentionally marked by the touch of the hand on the screen). Manual responses of drawing tests (Cube drawing, Clock drawing, ROCF) were recorded automatically and could be replayed. The scoring of drawing tests was done afterward by the neuropsychologist.

Semi-structured interview on user-experience

At the end of the test assessment, the neuropsychologist conducted a semi-structured interview consisting of eight questions: (1) What do you think about performing the tests on a tablet?; (2) How was the visibility of the tests?; (3) How difficult was drawing on a tablet screen?; (4a) How comparable was drawing on a tablet screen with drawing on paper?; (4b) What were the differences between drawing on a tablet screen and drawing on paper?; (5) Could you draw as precisely on a tablet screen as on paper?; (6) How accurate was the appearance of your drawing on the tablet screen?; (7) Was there a touch latency between the moment you drew and the appearance of your drawing on the tablet screen?; and (8) What improvements can be made? Response options ranged from 1 (negative) to 5 (positive) with different labels for each question, except for question 7, which could be answered with "yes" or "no". Question 4b and 8 were open-ended questions.

Demographic and clinical characteristics

We collected data on sex, age and level of education. Level of education was scored according to a Dutch classification system (Verhage, 1965), consisting of 7 levels, with 1 being the lowest (less than primary school) and 7 being the highest (academic degree). These levels were converted into three categories for analysis: low (Verhage 1–4), average (Verhage 5), and high (Verhage 6–7). This classification system is the most commonly used system in the Netherlands and is similar to the International Standard Classification of Education (UNESCO, 2012). We asked participants whether they used a tablet regularly, and, if yes, how many hours per week they used it. At the beginning of the test session, the conventional Mini-Mental State Examination – 2nd Edition (MMSE-2) was administered as measure of general cognitive functioning (Folstein et al., 2010). For stroke patients, we extracted time since stroke, stroke type (ischemic, hemorrhage or subarachnoid hemorrhage) and lesion side (left, right or both) from the medical files. For TBI patients, we extracted the time since injury, CT abnormalities (yes/no), and cause of injury (collision, fall, or other) from the medical files.

Statistical analysis

Demographic and clinical characteristics

Non-parametric tests (Kruskal-Wallis non-parametric ANOVA and post-hoc Mann-Whitney U test for continuous variables, and Chi-square test for categorical variables) were used to compare the demographic characteristics, tablet use, and global cognitive functioning between groups.

Feasibility and user-experience

To evaluate feasibility, we reported the number of stroke and TBI patients: (1) who did not complete one or more tests; (2) who needed more than one break during the test session; (3) for whom the brightness of the tablet screen had to be brought down to 50%; and (4) for whom the volume of the laptop needed to be turned down. To evaluate user-experience, we reported the responses for each closed-ended question of the semi-structured interview, split for stroke patients, TBI patients and healthy controls. For the open-ended questions, we described the answers that were provided by ≥5% of the participants.

Applicability of conventional norms on digital tests

Dutch conventional norms were applied to the raw scores of each outcome measure (See Supplementary Table 2).[1] The percentages of healthy controls, stroke patients and TBI patients who performed below the 10th percentile or below cutoff (RAVLT recognition, Cube Drawing, Clock Drawing, O-cancellation, Star Cancellation) were reported. Based on Lezak's distribution, we expected that <10% of the healthy controls would perform below the 10th percentile (Lezak et al., 2012). Regarding the stroke and TBI patients, we expected that >10% would perform below the 10th percentile, because of the expected cognitive disorders in these populations.

Effect of tablet familiarity on test performance

Based on the data of healthy controls, multiple linear regression analyses were conducted to explore the effect of tablet familiarity on test performance on each test of the d-NPA. The raw scores of the tests were used as outcome variables. We chose a hierarchical method (blockwise entry) where predictors were grouped into blocks. Age (in years), sex (coded as 0 [men] and 1 [women]) and level of education (dummy coded with average education as reference category) were used as predictors in the first block of the hierarchy (model 1). Tablet familiarity (use of tablet in hours per week) was added to the second block of the hierarchy (model 2). We evaluated the improvement of model 2 compared to model 1 by looking at the F-change and whether this change was significant. A Benjamini-Hochberg correction was applied to counteract the problem of multiple comparisons (Benjamini & Hochberg, 1995), which is considered the best approach in exploratory research (false discovery rate was set at .1).

Several assumptions were evaluated as followed: (1) multicollinearity between predictors was examined by inspecting Pearson's correlation coefficient (no significant correlations >.7); (2) independence of observations was evaluated by Durbin–Watson tests (values below 1 and above 3 are cause for concern); (3) the linearity and homoscedasticity were examined using scatter plots of residuals; (4) normality of residuals was examined by using probability-probability (p-p) plots; and (5) influential cases were identified by computing Cook's distances.

Results

Demographic and clinical characteristics

We invited 498 patients, of whom 378 patients did not respond or declined due to several reasons (e.g., no time/interest, personal reasons). We included 59 stroke patients and 61 TBI patients in our study. In addition, we included 56 healthy controls. We obtained d-NPA data of 103 healthy controls (from Philips Research), resulting in a total of 159 healthy controls. See Table 1 for the demographic and clinical characteristics.

The healthy control group and patient groups were comparable regarding the distribution of sex, education, handedness, global cognitive functioning and tablet use (% yes). There were no significant differences in the average number of hours they used a tablet per week (Table 1). Healthy controls were significantly older compared to stroke patients ($U = 3587.50$, $z = -2.67$, $p = .008$) and TBI patients ($U = 2688.00$, $z = -5.12$, $p < .001$).

Feasibility

The majority of the patients (94%) was able to complete the entire d-NPA (Table 2). One stroke patient was not able to complete the ROCF and not able to start the Stroop and the WCST, as the patient reported to be too tired. One TBI patient did not complete four tests (i.e., TMT, O-Cancellation, Star Cancellation, Stroop) due to sensory overload caused by the high density of stimuli (as reported by the patient). Of the five TBI patients who did not complete 1 to 2 tests, three patients additionally needed a reduction of the brightness, an adjustment of the volume, and/or an extra break. Of all patients, 5% needed an extra break and 6% needed technological adjustments.

User-experience

The majority of the participants (91%) considered performing the tests on a tablet as pleasant or very pleasant (Figure 1; question 1). Four patients reported the experience as (very) unpleasant, of which one TBI patient aborted four tests and one TBI patient aborted one test and needed an extra break and a reduction of the brightness. These patients reported that the unpleasant experience was caused by the brightness of the tablet screen which resulted in sensory overload (e.g., they felt it was

Table 1. Demographic and clinical characteristics.

	Patients with stroke $n = 59$	Patients with TBI $n = 61$	Healthy controls $n = 159$	Statistics
Male (%)	64.4	50.8	59.1	$\chi2(2) = 2.35, p = .308$
Age in years (mean, SD)	54.02 (13.26)	46.48 (16.21)	58.38 (13.82)	$H(2) = 28.31, p < .001$
Range	22–79	18–73	21–81	
Level of Education (%)				$\chi2(4) = 2.98, p = .561$
Low	13.6	9.8	8.2	
Average	27.1	21.3	20.8	
High	59.3	68.9	71.1	
Handedness* (%)				$\chi2(4) = 5.31, p = .257$
Left	15.3	13.1	6.3	
Right	83.1	83.6	91.2	
Ambidextrous	1.7	3.3	2.5	
Tablet use (% yes)	67.8	65.6	65.4	$\chi2(2) = .11, p = .945$
Tablet use hours per week (mean, SD)	5.29 (7.91)	6.29 (8.79)	5.49 (8.31)	$H(2) = .075, p = .963$
MMSE-2 (0–30) (mean, SD)	28.32 (1.96)	28.82 (1.37)	28.67 (1.43)	$H(2) = 1.66, p = .437$
Below cutoff of 24 (%)	1	1	1	
Time since stroke/TBI (mean)	19.4 months	45.8 months		
Range	(4–268)	(5–386)		
Stroke type (%)				
Ischemic	47.5			
Hemorrhage	8.5			
Subarachnoid hemorrhage	44.1			
Lesion side (%)				
Left	33.9			
Right	33.9			
Bilateral	15.3			
Not visible on scan	16.9			
CT abnormalities (%)				
Yes		54.1		
No		21.3		
Unknown		24.6		
Cause of injury (%)				
Collision		49.2		
Fall		41.0		
Other		9.8		
Testing site (%)				
University Medical Center	73	48	19	
Rehabilitation Center	7	31	4	
Participant's home	20	21	12	
Data obtained from Philips			65	

Abbreviations: Acquired Brain Injury (ABI), Computer Tomography (CT), Mini-Mental State Examination – 2nd version (MMSE-2), Standard Deviation (SD), Traumatic Brain Injury (TBI). *All patients used their dominant hand to perform the tests on the tablet, yet two patients whose injury affected the dominant hand used their non-dominant or both hands alternating.

Table 2. Feasibility of a digital administration of a neuropsychological assessment in stroke and TBI patients.

|	Patients with stroke $n = 59$	Patients with TBI $n = 61$
Completion d-NPA (%)	98.3 ($n = 58$)	90.2 ($n = 55$)
1–2 tests not completed (%)	0	8.2 ($n = 5$)
>2 tests not completed (%)	1.7 ($n = 1$)	1.6 ($n = 1$)
Need for extra break (%)	0	9.8 ($n = 6$)
Reduced brightness (%)	0	8.2 ($n = 5$)
Lowered sound volume (%)	1.7 ($n = 1$)	1.6 ($n = 1$)

Abbreviations: digital neuropsychological assessment (d-NPA), Traumatic Brain Injury (TBI).

Note. These observational measures were systematically administered in 56 healthy controls only. All of the healthy controls completed the d-NPA, and none of them needed an extra break, an adjustment of the brightness or sound volume.

tiring, required more mental energy). The visibility of the tests (question 2), the difficulty of drawing (question 3), and the appearance of the drawing (question 6) were considered satisfactory for patients and healthy controls. The majority of the participants (91%) reported there was no touch latency between the moment the participant drew and the appearance of the drawing on the tablet screen.

Different responses were provided regarding the precision of drawing on a tablet screen, with patients being more positive than healthy controls (question 5). Most patients and healthy controls reported that drawing on a tablet screen was quite similar with drawing on paper (question 4), however, there were noteworthy differences: the surface of the tablet screen gave less friction compared to drawing on paper (47%); drawing on a tablet screen was less accurate compared to drawing on paper (18%); errors could not be erased on the tablet (12%); one was not able to rest his/her hand on the tablet (9%); different manual feedback (e.g., the surface of the tablet felt "more distant" compared to paper) (5%); and the hand position was different when using a pencil stylus and tablet (5%).
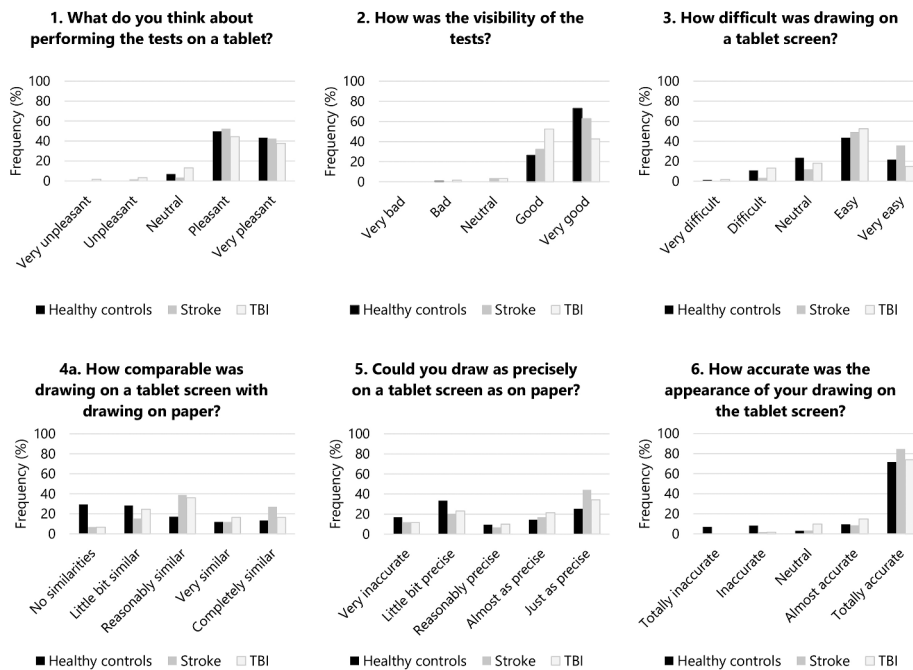
Figure 1. The six close-ended questions from the semi-structured interview are presented with the response options ranging from 1 (negative) to 5 (positive) with different labels for each question. The response options are presented on the horizontal axis. The frequency (%) of the reported response option is presented on the vertical axis, split per group.

Patients and healthy controls suggested the following improvements to the digital administration: increasing the degree of friction of the surface of the screen or the pencil stylus (8%); adjusting the brightness of the tablet screen to individual needs (5%); and improving the quality of the audio fragments (5%) (e.g., to announce the start of a test to get used to the monotonous computerized voice, use human speech). Two-thirds of the participants (67%) was satisfied with the digital administration and did not suggest any improvements.

Applicability of conventional norms on digital tests

Three stroke patients had been assessed with a conventional NPA in the three months prior to participation and were excluded for these analyses to prevent potential practice effects influencing the current results (Calamia et al., 2012). Table 3 shows the percentages of stroke patients, TBI patients and healthy controls showing an abnormal performance (<10th percentile or below cutoff) on each outcome measure (see Supplementary Table 3 for the average test scores and standard deviations per group). As expected, higher percentages of stroke and TBI patients performed abnormal on the tests when compared to healthy controls. Against expectations, more than 10% of the healthy controls showed abnormal performances on the RAVLT (immediate recall, delayed recall, recognition), TMT A, Clock Drawing, Cube Drawing, ROCF copy, Verbal Fluency Letter, Verbal

Table 3. Percentages of patients and healthy controls showing an "abnormal performance" based on Dutch conventional norms. Abnormal performance was defined as <10th percentile or below cutoff for the RAVLT recognition, Cube Drawing, Clock Drawing, O-cancellation, Star Cancellation.

Outcome measures	Patients with stroke n = 56 %	n	Patients with TBI n = 61 %	n	Healthy controls n = 159 %	n
RAVLT Immediate recall	44.6	56	41.7	60	**33.8***	157
RAVLT Delayed recall	35.7	56	25.0	60	**22.9***	157
RAVLT Delayed recall corrected	7.1	56	11.7	60	6.4	157
RAVLT Recognition	12.5	56	16.7	60	**11.4***	157
TMT A	42.9	56	40.0	60	**24.5***	159
TMT B	19.6	56	26.7	60	3.1	159
TMT A-B	8.9	56	15.0	60	2.5	159
Clock Drawing	35.7	56	29.5	61	**25.8***	159
Cube Drawing	26.8	56	31.1	61	**22.6***	159
O-Cancellation	5.4	56	0.0	60	3.8	159
Star Cancellation	1.8	56	6.7	60	6.9	159
ROCF Copy	30.4	56	34.4	61	**16.4***	159
ROCF Immediate recall	12.7	55	18.0	61	8.8	159
ROCF Delayed recall	14.5	55	18.0	61	9.4	159
Verbal Fluency Letter	25.0	56	36.1	61	**15.1***	159
Verbal Fluency Animals	17.9	56	23.0	61	6.3	159
Verbal Fluency Professions	28.6	56	23.0	61	**10.1***	159
Digital Span	16.1	56	37.7	61	8.8	159
WCST Total errors	16.4	55	6.8	59	7.6	157
WCST Perseverative errors	9.1	55	6.8	59	4.5	157
WCST Non-perseverative errors	14.5	55	6.8	59	9.6	157
WCST Number of completed categories	16.4	55	16.9	59	**12.7***	157
WCST Failure to maintain set	22.2	54	18.6	59	**18.6***	156

Abbreviations: Rey Auditory Verbal Learning Test (RAVLT) Trail Making Test (TMT); Rey-Osterrieth Complex Figure (ROCF); Wisconsin Card Sorting Test (WCST). *More than 10% of the participants performing below 10th percentile or below cutoff (depicted **in bold**).

Fluency Professions, WCST number of completed categories, and the WCST failure to maintain set.

Effect of tablet familiarity on test performance

With regard to the assumptions, no multicollinearity was examined, there was independence of observations, and no influential cases were identified. The scatter plots demonstrated linear relationships between the dependent and independent variables and homoscedasticity, except for the O-Cancellation, Star Cancellation, ROCF copy, WCST number of completed categories, and WCST failure to maintain set (see Supplementary Figure 1ab). The p-p plots showed normally distributed

standardized residuals, except for the O-Cancellation, Star Cancellation, Stroop 1, Stroop 2, WCST number of completed categories, and WCST failure to maintain set were cause for concern (see Supplementary Figure 2ab).

Significant effects of age, sex, and level of education (model 1) were found on each outcome measure of the digital tests, except for the O-Cancellation (Table 4). There was no significant improvement in predicting the outcome measures of the digital tests when adding technological familiarity as new predictor (model 2). This finding suggests there was no significant effect of tablet familiarity on test performance on any of the outcome measures of the d-NPA.

Table 4. Results of the multiple regression analyses by using a hierarchical method based on the data of healthy controls.

Outcome measures	Model	R^2	F-change	Sig F-change	n
RAVLT Immediate recall	Age, sex, education	.44	30.08	<.001	158
	Age, sex, education, tablet familiarity	.45	1.20	**.276**	
RAVLT Delayed recall	Age, sex, education	.39	24.66	<.001	158
	Age, sex, education, tablet familiarity	.40	0.73	**.394**	
RAVLT Recognition	Age, sex, education	.19	9.17	<.001	158
	Age, sex, education, tablet familiarity	.19	0.13	**.716**	
TMT A	Age, sex, education	.29	15.54	<.001	159
	Age, sex, education, tablet familiarity	.29	0.03	**.872**	
TMT B	Age, sex, education	.32	17.93	<.001	159
	Age, sex, education, tablet familiarity	.32	0.29	**.594**	
O-Cancellation	Age, sex, education	.04	1.38	.242	159
	Age, sex, education, tablet familiarity	.04	0.01	**.909**	
Star Cancellation	Age, sex, education	.08	3.10	.017	159
	Age, sex, education, tablet familiarity	.10	4.07	**.046**	
ROCF Copy	Age, sex, education	.14	6.45	<.001	159
	Age, sex, education, tablet familiarity	.17	5.13	**.025**	
ROCF Immediate recall	Age, sex, education	.21	9.92	<.001	159
	Age, sex, education, tablet familiarity	.22	2.23	**.137**	
ROCF Delayed recall	Age, sex, education	.26	13.82	<.001	159
	Age, sex, education, tablet familiarity	.27	1.17	**.282**	
Verbal Fluency Letter	Age, sex, education	.16	7.49	<.001	159
	Age, sex, education, tablet familiarity	.16	0.10	**.751**	
Verbal Fluency Animals	Age, sex, education	.16	7.40	<.001	159
	Age, sex, education, tablet familiarity	.17	0.67	**.413**	
Verbal Fluency Professions	Age, sex, education	.05	2.16	.076	159
	Age, sex, education, tablet familiarity	.05	0.14	**.714**	
Verbal Fluency Fruit/furniture	Age, sex, education	.20	9.76	<.001	159
	Age, sex, education, tablet familiarity	.20	0.03	**.871**	
Digit Span	Age, sex, education	.21	10.12	<.001	159
	Age, sex, education, tablet familiarity	.21	0.50	**.479**	
Stroop 1	Age, sex, education	.09	3.64	.007	156
	Age, sex, education, tablet familiarity	.09	0.00	**.992**	
Stroop 2	Age, sex, education	.12	4.89	.001	154
	Age, sex, education, tablet familiarity	.12	0.01	**.941**	
Stroop 3	Age, sex, education	.24	11.81	<.001	153
	Age, sex, education, tablet familiarity	.24	0.39	**.532**	
WCST Total errors	Age, sex, education	.20	9.66	<.001	157
	Age, sex, education, tablet familiarity	.20	0.31	**.862**	
WCST Perseverative errors	Age, sex, education	.17	7.87	<.001	157
	Age, sex, education, tablet familiarity	.17	0.11	**.916**	
WCST Non-perseverative errors	Age, sex, education	.20	9.53	<.001	157
	Age, sex, education, tablet familiarity	.20	0.06	**.81**	
WCST Number of completed categories	Age, sex, education	.18	8.13	<.001	157
	Age, sex, education, tablet familiarity	.18	0.11	**.737**	
WCST Failure to maintain set	Age, sex, education	.11	4.65	<.001	157
	Age, sex, education, tablet familiarity	.11	0.36	**.551**	

Note. *F-change* represents the improvement in predicting the outcome measure by adding a new predictor to the model. We evaluated whether this change was significant *(Sig F-change*; **in bold**). Based on a Benjamini Hochberg correction, there was no significant improvement.

Discussion

In this study, we investigated (1) the feasibility and user-experience of a d-NPA in patients with ABI and healthy controls; (2) the applicability of conventional norms on digital tests, and (3) whether familiarity with a tablet would affect test performance on a tablet. We found that the administration of a d-NPA seems feasible for cognitive assessment in patients with ABI. The digital administration was considered a pleasant experience for patients with ABI and healthy controls. Only 6% of the patients was unable to complete the d-NPA, 5% needed an extra break, and 6% needed an adjustment of the brightness and/or volume. Patients who did not complete the d-NPA reported mental fatigue or sensory overload caused by an overdose of stimuli and/or the brightness of the tablet screen. As we did not directly compare the d-NPA with a conventional NPA, we cannot rule out the possibility that these patients would have experienced sensory overload with paper-and-pencil tests as well, as sensory overload may be caused by various factors (e.g., task demand, fatigue) (Scheydt et al., 2017). The brightness of the tablet screen, however, may add to the sensory overload and adjusting brightness might be a proper solution to suit individual needs. However, brightness and/or luminance contrast can have an impact on the readability or visibility of visual stimuli (Schlegel & Gilliland, 2007), so future research should investigate how adjustments in brightness and contrast impacts test performance and develop adapted norms for brightness and/or contrast levels, when this may affect performance.

The conventional paper-and-pencil norms were not applicable for half of the digital tests, as up to 34% of healthy controls showed an abnormal performance (<10th percentile or below cutoff) (Lezak et al., 2012). There are several possible explanations for this result. An explanation may be the subtle – but relevant – differences in administration (paper-and-pencil vs. tablet-and-pencil stylus) that might have influenced test performance. For instance, patients and healthy controls reported that the tablet screen gave less friction when drawing with the pencil stylus. Due to low friction, people tend to draw faster on a tablet than with pencil on a paper (Gerth et al., 2016; Guilbert et al., 2019), which might result in an unprecise drawing (see Figure 2 with the ROCF as an example). Furthermore, the quality of the speech synthesizer (i.e., artificial production of human speech) may have influenced the clarity. In especially the RAVLT, it may therefore have been difficult to correctly identify the words. Finally, changes in the nature of a response and feedback may also affect test performances (Schlegel & Gilliland, 2007). For instance, in the WCST, virtual cards were displayed on the tablet (instead of the use of real cards), and the participant received written feedback (instead of verbal feedback). Previous studies reported that normative data that exists for paper-and-pencil tests cannot simply be applied to digital tests, as performances on paper-and-pencil and digital tests are not directly
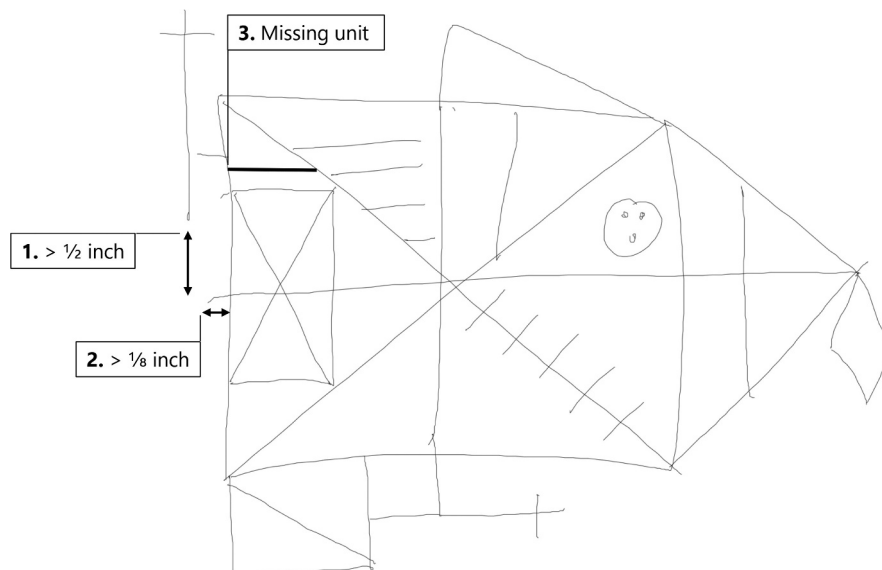
Figure 2. Example of a copy of the Rey-Osterrieth Complex Figure performed by a healthy control (30 years of age) with a total score of 32. Even though most units were present, unprecise drawing (highlighted in the example) resulted in a weak performance. For instance, (1) the height of the vertical cross should not extend more than ½ inch above the horizontal line (minus 1 point); (2) The horizontal line should not overshoot the vertical segments of the large rectangle more than ⅛ inch (minus 1 point); and (3) a horizontal line should be drawn parallel to and directly above the small rectangle (minus 2 points) (Meyers & Meyers, 1995).

comparable (Bauer et al., 2012; Germine et al., 2019; Parsey & Schmitter-Edgecombe, 2013). For this reason, even when a digital test mirrors a paper-and-pencil test, new clinical norms are needed (Bauer et al., 2012).

Another important factor might regard the characteristics of the conventional norms used in this study. Norms are ideally updated regularly (Germine et al., 2019). However, many paper-and-pencil tests exist for decades and test performances are interpreted using norms from studies that were conducted several decades ago (in this study ranging from the year 1993–2012) (Bilder & Reise, 2019; Dickinson & Hiscock, 2011). General experiences of a population change over time and highly affect test performance, also known as the Flynn effect (Dickinson & Hiscock, 2011). The Flynn effect refers to the rise of scores on intelligence and neuropsychological tests throughout the 20th century. In contrast to the Flynn effect, a surprisingly high number (26%) of healthy participants were not able to draw a clock correctly. Participants placed the numbers outside the contour or even placed the hands incorrectly, which is unlikely a result of differences in the means of assessment (paper-and-pencil vs. tablet-and-pencil stylus). Environmental changes resulting from modernization – such as greater use of technology – might result in the fact that people are increasingly accustomed to digital clocks. Previous studies have described concerns regarding the long-term use of the clock-drawing test due to the advent of digital clocks (Hazan et al., 2017; Shulman, 2000). Furthermore, normative data that derived from a specific population may not be generalizable to different populations (Lezak et al., 2004), which may result in false positives or negatives (see Supplementary Table 2 for the normative population characteristics). In short, developing and regularly updating clinical norms is crucial in neuropsychological assessment (Dickinson & Hiscock, 2011; Germine et al., 2019) and should be taken into account in order to implement a d-NPA in clinical practice.

Results on previous studies suggested that people with greater computer experience tend to perform better on computerized tests than those with less computer experience (Iverson et al., 2009; Tun & Lachman, 2010). Here, familiarity with a tablet did not affect cognitive test performances. This finding is consistent with a recent study of Wallace et al. (2019), who also found no differences in test performances between TBI patients who reported to be less or more comfortable with an iPad (Wallace et al., 2019). Touch-screen devices require little training, little cognitive demands, and little hand-eye coordination and are therefore easy to use, even by individuals who are minimally exposed to technology (Canini et al., 2014; Cernich et al., 2007; Holzinger, 2010; Wood et al., 2005). Therefore, in d-NPA, tablets should be chosen over computers with keyboard, computer mouse or joystick.

Strength and limitations

An important strength of this study was the engagement of a large number of stroke and TBI patients ($n = 120$). The importance of including end users in the development and evaluation of new medical technological devices is more and more acknowledged and stressed (Jenkins et al., 2016; Shah & Robinson, 2007). We intentionally aimed for a large and heterogeneous sample of patients to increase its representativeness. A general concern might regard a potential selection bias, where patients who are willing to participate are probably patients who are less impaired (Knudsen et al., 2010; Olson et al., 2010). Our patient samples were relatively young and moderately impaired, which might be regarded as a limitation since we cannot generalize the current findings to an older and/ or more impaired population. We suspect that a d-NPA might be a bit more challenging with an older population and/or a more impaired population, just like the conventional NPA would be. As such, there is no indication that the d-NPA would not be feasible for other groups, yet this remains to be tested.

One potential limitation is that injury characteristics were not systematically noted in the medical files, and we were therefore unable to further investigate specific subgroups within the patient samples. For example, it would have been interesting to investigate whether lesion location, volume, severity of stroke or TBI determined by classification measures (e.g., the Glasgow Coma Scale, duration loss of consciousness or post-traumatic amnesia) would affect the feasibility or user-experience. With the current results, where the majority of the patients (94%) were able to complete the d-NPA and considered it as "pleasant", there is no direct reason to assume that there would be very large deviations within specific subgroups.

One might argue that the design of the study is not ideal, as we did not directly compare a paper-and-pencil and a digital administration. Even though we did not aim for a direct comparison, we do feel that we need to address this alternative design. When one would be interested in examining differences between a paper-and-pencil and a digital administration with respect to user-experience and test performances, participants would need to be assessed twice with the same tests. A long duration between sessions would be necessary as otherwise mainly practice effects would be assessed (Calamia et al., 2012). In general,

diminishing or removing practice effects is challenging when using a within-subjects design in neuropsychological research. For this reason, we investigated the feasibility and user-experience of a d-NPA, without the direct comparison with a conventional NPA.

Finally, there are two drawbacks in the current study. First, the average age of healthy controls was significantly higher compared to the patient groups. Conventional norms, however, correct for the effect of age on test performances, and as such the current results – corrected for age – still hold. Second, we used self-report to measure tablet familiarity, as we asked participants to estimate how many hours per week they use a tablet. It is generally accepted that the validity of retrospective self-reports may be compromised, due to for example, a limited autobiographical memory (Schwarz, 2007). Even though tablet familiarity did not seem to have an important effect on test performance, alternative measures would have possibly been more suitable, such as measures capturing real-time data (e.g., diaries, applications that register how much time people spend on a tablet).

Future research

Based on our findings, researchers and manufacturers should collaborate to reduce potential restrictions for optimal use (e.g., low friction of tablet screens, low quality of speech synthesizer) that interfere with the user-experience and usability of such devices. A d-NPA offers several advantages over a paper-and-pencil assessment, such as a more standardized administration with an increased accuracy and timing of stimulus presentation. A d-NPA allows an automatized scoring which saves valuable professional time and is less prone to human errors. In addition, manual and verbal responses can be replayed afterward, avoiding observations or order of responses to be lost. Moreover, a digital response acquisition allows for a highly precise and detailed data collection, which opens the possibility to develop novel outcome measures to assess subtle cognitive impairments (Davis et al., 2015; Diaz-Orueta et al., 2020; Parsey & Schmitter-Edgecombe, 2013). A next step should be the development of additional outcome measures that go beyond the traditional outcome measures of paper-and-pencil tests. Accurate time measures could reveal fluctuations in test performance (Spreij et al., n.d.), and algorithms could improve the assessment of the process of construction in drawing tests (Kim et al., 2011). Finally, the development of new norms remains crucial in order to implement a d-NPA in clinical practice (Germine et al., 2019).

Conclusions

The administration of a d-NPA is feasible in patients with ABI. The digital administration was considered a pleasant experience for patients with ABI and healthy controls. Familiarity with a tablet did not impact test performance, which is particularly important in neuropsychological assessment. Conventional norms that exist for the paper-and-pencil tests were not applicable on the digital version of the tests. Future research should focus on developing norms in order to implement a d-NPA in clinical practice.

Note

1. The Stroop was not included in these analyses. Dutch clinical norms of the Stroop are based on stimuli where subsequent colors are sometimes the same (e.g., red, red, red), whereas our digital version included stimuli where subsequent colors are never the same (Hammes, 1973). The clinical norms were therefore not applicable to our digital version of the test, as pronouncing the same color subsequently improves velocity.

Disclosure statement

No potential conflict of interest was reported by the author.

ORCID

Lauriane A. Spreij ⓘD http://orcid.org/0000-0001-5558-9746
Isabel K. Gosselt ⓘD http://orcid.org/0000-0001-8086-8843
Johanna M. A. Visser-Meily ⓘD http://orcid.org/0000-0002-5955-8012
Tanja C. W. Nijboer ⓘD http://orcid.org/0000-0001-6683-0267

References

American Psychological Association. (Ed.). (1986). *Guidelines for computer-based tests and interpretations.*

Baker, E., Letz, R. E., Fidler, A. T., Shalat, S., Plantamura, D., & Lyndon, M. (1985). A computer-based neurobehavioral evaluation system for occupational and environmental epidemiology: Methodology and validation studies. *Neurobehavioral Toxicology and Teratology*, *7*(4), 369–377.

Bauer, R. M., Iverson, G. L., Cernich, A. N., Binder, L. M., Ruff, R. M., & Naugle, R. I. (2012). Computerized neuropsychological assessment devices: joint position paper of the american academy of clinical neuropsychology and the national academy of neuropsychology. *Archives of Clinical Neuropsychology*, *27*(3), 362–373. https://doi.org/10.1093/arclin/acs027

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methological)*, *57*(1), 289–300. https://www.jstor.org/stable/2346101

Bilder, R. M., & Reise, S. P. (2019). Neuropsychological tests of the future: How do we get there from here? *The Clinical Neuropsychologist*, *33*(2), 220–245. https://doi.org/10.1080/13854046.2018.1521993

Calamia, M., Markon, K., & Tranel, D. (2012). Scoring higher the second time around: Meta-analyses of practice effects in neuropsychological assessment. *The Clinical Neuropsychologist*, *26*(4), 543–570. https://doi.org/10.1080/13854046.2012.680913

Canini, M., Battista, P., Della Rosa, P. A., Catricalà, E., Salvatore, C., Gilardi, M. C., & Castiglioni, I. (2014). Computerized neuropsychological assessment in aging: Testing efficacy and clinical ecology of different interfaces. *Computational and Mathematical Methods in Medicine*, 2014, 804723. https://doi.org/10.1155/2014/804723

Carr, A. C., Woods, R. T., & Moore, B. J. (1986). Automated cognitive assessment of elderly patients: A comparison of two types of response device. *British Journal of Clinical Psychology*, *25*(4), 305–306. https://doi.org/10.1111/j.2044-8260.1986.tb00711.x

Cernich, A. N., Brennana, D. M., Barker, L. M., & Bleiberg, J. (2007). Sources of error in computerized neuropsychological assessment. *Archives of Clinical Neuropsychology*, *22*(SUPPL. 1), 39–48. https://doi.org/10.1016/j.acn.2006.10.004

Commission, I. T. (2001). International guidelines for test use. *International Journal of Testing*, *1*(2), 93–114. https://doi.org/10.1207/S15327574IJT0102

Davis, R., Libon, D. J., Au, R., Pitman, D., & Penney, D. L. (2015). THink: Inferring cognitive status from subtle behaviors. *AI Magazine*, *36*(3), 49–60. https://doi.org/10.1609/aimag.v36i3.2602

Diaz-Orueta, U., Blanco-Campal, A., Lamar, M., Libon, D. J., & Burke, T. (2020). Marrying past and present neuropsychology: is the future of the process-based approach technology-based? *Frontiers in Psychology*, *11*(March), 1–15. https://doi.org/10.3389/fpsyg.2020.00361

Dickinson, M. D., & Hiscock, M. (2011). The Flynn effect in neuropsychological assessment. *Applied Neuropsychology*, *18*(2), 136–142. https://doi.org/10.1080/09084282.2010.547785

Elwood, R. W. (2001). MicroCog: Assessment of cognitive functioning. *Neuropsychology Review*, *11*(2), 89–100. https://doi.org/10.1023/A:1016671201211

Findlater, L., Froehlich, J. E., Fattal, K., Wobbrock, J. O., & Dastyar, T. (2013). Age-related differences in performance with touchscreens compared to traditional mouse input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)* (pp. 343–346). Association for Computing Machinery. https://doi.org/10.1145/2470654.2470703

Folstein, M. F., Folstein, S. E., White, T., & Messer, M. A. (2010). *MMSE-2: Mini-mental state examination* (2nd ed.). Psychological Assessment Resources, Inc.

Germine, L., Reinecke, K., & Chaytor, N. S. (2019). Digital neuropsychology: Challenges and opportunities at the intersection of science and software. *Clinical Neuropsychologist*, *33*(2), 271–286. https://doi.org/10.1080/13854046.2018.1535662

Gershon, R. C., Cella, D., Fox, N. A., Havlik, R. J., Hendrie, H. C., & Wagster, M. V. (2010). Assessment of neurological and behavioural function: The NIH Toolbox. *The Lancet Neurology*, *9*(2), 138–139. https://doi.org/10.1016/S1474-4422(09)70335-7

Gerth, S., Dolk, T., Klassert, A., Fliesser, M., Fischer, M. H., Nottbusch, G., & Festman, J. (2016). Adapting to the surface: A comparison of handwriting measures when writing on a tablet computer and on paper. *Human Movement Science*, *48*, 62–73. https://doi.org/10.1016/j.humov.2016.04.006

Gualtieri, C. T., & Johnson, L. G. (2006). Reliability and validity of a computerized neurocognitive test battery, CNS Vital Signs. *Archives of Clinical Neuropsychology*, *21*(7), 623–643. https://doi.org/10.1016/j.acn.2006.05.007

Guilbert, J., Alamargot, D., & Morin, M. F. (2019). Handwriting on a tablet screen: Role of visual and proprioceptive feedback in the control of movement by children and adults. *Human Movement Science*, *65*(September), 30–41. https://doi.org/10.1016/j.humov.2018.09.001

Hammes, J. (1973). *De Stroop Kleur-Woord Test: Handleiding [The Stroop Color-Word Test: Manual]*. Swets & Zeitlinger.

Harvey, P. D. (2012). Clinical applications of neuropsychological assessment. *Dialogues in Clinical Neuroscience*, *14*(1), 91–99.

Hazan, E., Frankenburg, F., Brenkel, M., & Shulman, K. (2017). The test of time: A history of clock drawing. *International Journal of Geriatric Psychiatry*, *33*(1), e22–e30. https://doi.org/10.1002/gps.4731

Heaton, R. K., & Matthews, C. G. (1986). Differences in neuropsychological test performance associated with age, education, and sex. In I. Grant & K. M. Adams (Eds.), *Neuropsychological assessment in neuropsychiatric disorders: Clinical methods and empirical findings* (pp. 100–120). Oxford University Press.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin card sorting test manual revised and expanded* (2nd ed.). Psychological Assessment Resources, Inc.

Holzinger, A. (2010). Finger instead of mouse: touch screens as a means of enhancing universal access. IIn N. Carbonell, & C. Stephanidis (Eds.), *Universal access : Theoretical perspectives, practice, and experience* (Vol. 2615, pp. 387–397). (Lecture notes in computer sciences). Springer. https://doi.org/10.1007/3-540-36572-9_30

Iverson, G. L., Brooks, B. L., Ashton, V. L., Johnson, L. G., & Gualtieri, C. T. (2009). Does familiarity with computers affect computerized neuropsychological test performance? *Journal of Clinical and Experimental Neuropsychology*, *31*(5), 594–604. https://doi.org/10.1080/13803390802372125

Jenkins, A., Lindsay, S., Eslambolchilar, P., Thornton, I. M., & Tales, A. (2016). Administering cognitive tests through touch screen tablet devices: Potential issues. *Journal of Alzheimer's Disease*, *54*(3), 1169–1182. https://doi.org/10.3233/JAD-160545

Joddrell, P., & Astell, A. J. (2016). Studies involving people with dementia and touchscreen technology: A literature review. *JMIR Rehabilitation and Assistive Technologies*, *3*(2), e10. https://doi.org/10.2196/rehab.5788

Kabat, M. H., Kane, R. L., Jefferson, A. L., & DiPino, R. K. (2001). Construct validity of selected automated neuropsychological assessment metrics (ANAM) battery measures. *Clinical Neuropsychologist*, *15*(4), 498–507. https://doi.org/10.1076/clin.15.4.498.1882

Kane, R. L., & Kay, G. G. (1992). Computerized assessment in neuropsychology: A review of tests and test batteries. *Neuropsychology Review*, *3*(1), 1–117. https://doi.org/10.1007/BF01108787

Kane, R. L., & Reeves, D. L. (1997). *The neuropsychology handbook: Automated neuropsychological assessment metrics (ANAM)*. Springer.

Kaplan, E. (1988). The process approach to neuropsychological assessment. *Aphasiology*, *2*(3–4), 309–311. https://doi.org/10.1080/02687038808248930

Kim, H., Cho, Y. S., & Do, E. Y. L. (2011). Computational clock drawing analysis for cognitive impairment screening. *Proceedings of the 5th International Conference on Tangible Embedded and Embodied Interaction*, *TEI'11*(May 2014), 297–300. https://doi.org/10.1145/1935701.1935768

Knudsen, A. K., Hotopf, M., Skogen, J. C., Øverland, S., & Mykletun, A. (2010). The health status of nonparticipants in a population-based health study. *American Journal of Epidemiology*, *172*(11), 1306–1314. https://doi.org/10.1093/aje/kwq257

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment (5th ed.)*. Oxford University Press.

Lezak, M. D., Howieson, D. B., Loring, D. W., Hannay, H. J., & Fischer, J. S. (2004). *Neuropsychological assessment* (4th ed.). Oxford University Press.

Maroon, J. C., Lovell, M. R., Norwig, J., Podell, K., Powell, J. W., & Hartl, R. (2000). Cerebral concussion in athletes: Evaluation and neuropsychological testing. *Neurosurgery*, *47*(3), 659–672. https://doi.org/10.1227/00006123-200009000-00027

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

Muñiz, J., & Bartram, D. (2007). Improving international tests and testing. *European Psychologist*, *12*(3), 206–219. https://doi.org/10.1027/1016-9040.12.3.206

Murata, A., & Iwase, M. (2005). Usability of touch-panel interfaces for older adults. *Human Factors*, *47*(4), 767–776. https://doi.org/10.1518/001872005775570952

Nasreddine, Z. S., Philips, N. A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., . . . Chertkow, H. (2005). The montreal cognitive assessment, MoCA : A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Nyhus, E., & Barceló, F. (2009). The Wisconsin Card Sorting Test and the cognitive assessment of prefrontal executive functions : A critical update. *Brain and Cognition*, *71*(3), 437–451. https://doi.org/10.1016/j.bandc.2009.03.005

Olson, R., Parkinson, M., & McKenzie, M. (2010). Selection bias introduced by neuropsychological assessments. *Canadian Journal of Neurological Sciences*, *37*(2), 264–268. https://doi.org/10.1017/S0317167100010039

Parsey, C. M., & Schmitter-Edgecombe, M. (2013). Applications of technology in neuropsychological assessment. *Clinical Neuropsychologist*, *27*(8), 1328–1361. https://doi.org/10.1080/13854046.2013.834971

Paul, R. H., Lawrence, J., Williams, L. M., Richard, C. C., Cooper, N., & Gordon, E. (2005). Preliminary validity of "integneuro™": A new computerized battery of neurocognitive tests. *International Journal of Neuroscience*, *115*(11), 1549–1567. https://doi.org/10.1080/00207450590957890

Powell, D. H., Kaplan, E. F., Whitla, D., Catlin, R., & Funkenstein, H. H. (1993). *Microcog: Assessment of cognitive functioning version 2.1*. The Psychological Corporation.

Rabin, L. A., Spadaccini, A. T., Brodale, D. L., Grant, K. S., Elbulok-Charcape, M. M., & Barr, W. B. (2014). Utilization rates of computerized tests and test batteries among clinical neuropsychologists in the United States and Canada. *Professional Psychology, Research and Practice*, *45*(5), 368–377. https://doi.org/10.1037/a0037987

Rasquin, S., Ooms, N., van de Sande, P., Beers, K., & Schmand, B. (2009). Validity and data of reference of a visual-spatial search task: O-search-task [in Dutch]. *Tijdschrift Voor Neuropsychologie*, *3*, 44–54.

Reitan, R. M. (1955). The relation of the trail making test to organic brain damage. *Journal of Consulting Psychology*, *19*(5), 393–394. https://doi.org/10.1037/h0044509

Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., Lawrence, A. D., Mcinnes, L., & Rabbitt, P. M. A. (1998). A study of performance on tests from the CANTAB battery sensitive to frontal lobe dysfunction in a large sample of normal volunteers: Implications for theories of executive functioning and cognitive aging. *Journal of the International Neuropsychological Society*, *4*(5), 474–490. https://doi.org/10.1017/S1355617798455073

Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., McInnes, L., & Rabbitt, P. (1994). Cambridge neuropsychological test automated battery (CANTAB): A factor analytic study of a large sample of normal elderly volunteers. *Dementia*, *5*(5), 266–281. https://doi.org/10.1159/000106735

Saan, R. J., & Deelman, B. G. (1986). *De 15-Woordentest A en B. (Een voorlopige handleiding)*. Afdeling Neuropsychologie, AZG (interne publicatie).

Scheydt, S., Müller Staub, M., Frauenfelder, F., Nielsen, G. H., Behrens, J., & Needham, I. (2017). Sensory overload: A concept analysis. *International Journal of Mental Health Nursing*, *26*(2), 110–120. https://doi.org/10.1111/inm.12303

Schlegel, R. E., & Gilliland, K. (2007). Development and quality assurance of computer-based assessment batteries. *Archives of Clinical Neuropsychology*, *22*(SUPPL. 1), 49–61. https://doi.org/10.1016/j.acn.2006.10.005

Schmand, B., Groenink, S. C., & Van Den Dungen, M. (2008). Letterfluency: Psychometrische eigenschappen en Nederlandse normen (Letter fluency: Psychometric properties and Dutch normative data; Article in Dutch). *Tijdschrift Voor Gerontologie En Geriatrie*, *39*(2), 64–76. https://doi.org/10.1007/bf03078128

Schmand, B., Houxen, P., de Koning, I., Gerritsen, M., Hoogman, M., Muslimovic, D., . . . van Tricht, M. *Normen neuropsychologische tests 2012 [Dutch Normative Data Neuropsychological Tests 2012].*, Pub. L. No. 2012, Nederlands Instituut van Psychologen (2012).

Schwarz, N. (2007). Retrospective and concurrent self-reports: The rationale for real-time data capture. In *The science of real-time data capture: Self-reports in health research* (pp. 11–26). Oxford University Press.

Shah, S. G. S., & Robinson, I. (2007). Benefits of and barriers to involving users in medical device technology development and evaluation. *International Journal of Technology Assessment in Health Care*, 23(1), 131–137. https://doi.org/10.1017/S0266462307051677

Shuhfried, G. (2013). *Vienna test system: Psychological assessment.* Shuhfried.

Shulman, K. I. (2000). Clock-drawing: Is it the ideal cognitive screening test? *International Journal of Geriatric Psychiatry*, 15(6), 548–561. https://doi.org/10.1002/1099-1166(200006)15:6<548::AID-GPS242>3.0.CO;2-U

Silverstein, S. M., Berten, S., Olson, P., Paul, R., Williams, L. M., Cooper, N., & Gordon, E. (2007). Development and validation of a World-Wide-Web-based neurocognitive assessment battery: WebNeuro. *Behavior Research Methods*, 39(4), 940–949. https://doi.org/10.3758/BF03192989

Spreij, L. A., Gosselt, I. K., Visser-Meily, J. M. A., Hoogerbrugge, A., Kootstra, T., & Nijboer, T. C. W. (n. d.). *The journey is just as important as the destination - digital neuropsychological assessment provides performance stability measures.*

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd.). Oxford University Press.

Tun, P. A., & Lachman, M. E. (2010). The association between computer use and cognition across adulthood: Use it so you won't lose it? *Psychological Aging*, 25(3), 560–568. https://doi.org/10.1037/a0019543

UNESCO Institute for Statistics. (2012). *International standard classification of education: ISCED 2011.*

Van der Stoep, N., Visser-Meily, J. M. A., Kappelle, L. J., de Kort, P. L. M., Huisman, K. D., Eijsackers, A. L. H., . . . Nijboer, T. C. W. (2013). Exploring near and far regions of space: Distance-specific visuospatial neglect after stroke. *Journal of Clinical and Experimental Neuropsychology*, 35 (8), 799–811. https://doi.org/10.1080/13803395.2013.824555

Verhage, F. (1965). Intelligence and age in a dutch sample. *Human Development*, 8(4), 238–245. https://doi.org/10.1159/000270308

Vermeent, S., Dotsch, R., Schmand, B., Klaming, L., Miller, J. B., & van Elswijk, G. (2020). Evidence of validity for a newly developed digital cognitive test battery. *Frontiers in Psychology*, 11(April), 1–11. https://doi.org/10.3389/fpsyg.2020.00770

Wallace, S. E., Donoso Brown, E. V., Schreiber, J. B., Diehl, S., Kinney, J., & Zangara, L. (2019). Touchscreen tablet-based cognitive assessment versus paper-based assessments for traumatic brain injury. *NeuroRehabilitation*, 45(1), 25–36. https://doi.org/10.3233/NRE-192725

Wechsler, D. (2004). *WAIS-III Nederlandse bewerking – Derde editie. Afname en scoringshandleiding.* Harcourt Test Publishers.

Wilson, B., Cockburn, J., & Halligan, P. W. (1987). Development of a behavioral test of visuospatial neglect. *Archives of Physical Medicine and Rehabilitation*, 68(2), 98–102.

Witt, J. A., Alpherts, W., & Helmstaedter, C. (2013). Computerized neuropsychological testing in epilepsy: Overview of available tools. *Seizure*, 22(6), 416–423. https://doi.org/10.1016/j.seizure.2013.04.004

Wood, E., Willoughby, T., Rushing, A., Bechtel, L., & Gilbert, J. (2005). Use of computer input devices by older adults. *Journal of Applied Gerontology*, 24(5), 419–438. https://doi.org/10.1177/0733464805278378