

Initial validation of a web-based self-administered neuropsychological test battery for older adults and seniors

Tor Ivar Hansen^{1,2}, Elise Christina D. Haferstrom², Jan F. Brunner^{1,3}, Hanne Lehn^{2,4}, and Asta Kristine Håberg^{1,2}

¹Department of Neuroscience, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

²Department of Medical Imaging, St. Olavs University Hospital, Trondheim, Norway

³Department of Physical Medicine and Rehabilitation, St. Olavs University Hospital, Trondheim, Norway

⁴Faculty of Medicine, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

(Received 7 August 2014; accepted 30 March 2015)

Introduction: Computerized neuropsychological tests are effective in assessing different cognitive domains, but are often limited by the need of proprietary hardware and technical staff. Web-based tests can be more accessible and flexible. We aimed to investigate validity, effects of computer familiarity, education, and age, and the feasibility of a new web-based self-administered neuropsychological test battery (Memoro) in older adults and seniors. **Method:** A total of 62 (37 female) participants (mean age 60.7 years) completed the Memoro web-based neuropsychological test battery and a traditional battery composed of similar tests intended to measure the same cognitive constructs. Participants were assessed on computer familiarity and how they experienced the two batteries. To properly test the factor structure of Memoro, an additional factor analysis in 218 individuals from the HUNT population was performed. **Results:** Comparing Memoro to traditional tests, we observed good concurrent validity ($r = .49-.63$). The performance on the traditional and Memoro test battery was consistent, but differences in raw scores were observed with higher scores on verbal memory and lower in spatial memory in Memoro. Factor analysis indicated two factors: verbal and spatial memory. There were no correlations between test performance and computer familiarity after adjustment for age or age and education. Subjects reported that they preferred web-based testing as it allowed them to set their own pace, and they did not feel scrutinized by an administrator. **Conclusions:** Memoro showed good concurrent validity compared to neuropsychological tests measuring similar cognitive constructs. Based on the current results, Memoro appears to be a tool that can be used to assess cognitive function in older and senior adults. Further work is necessary to ascertain its validity and reliability.

Keywords: Memory; Internet; Assessment; Computer; HUNT.

Computerized neuropsychological tests have become an integral part of neuropsychological practice both in research and in the clinic in the

past decade (Iverson, Brooks, Ashton, Johnson, & Gualtieri, 2009). With the aging population it is important to have access to valid and effective

The authors would like to thank the Nord-Trøndelag Health Study (the HUNT Study) for allowing us to do web-based cognitive testing in their cohort. The HUNT Study is a collaboration between HUNT Research Centre (Faculty of Medicine, Norwegian University of Science and Technology NTNU), Nord-Trøndelag County Council, Central Norway Health Authority, and the Norwegian Institute of Public Health.

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

Address correspondence to: Tor Ivar Hansen, Department of Neuroscience, Faculty of Medicine, Pb. 8905, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway (E-mail: torivar.hansen@ntnu.no).

methods to assess cognitive function in large cohorts, for example to detect subtle changes in individuals' cognitive abilities over time (Falletti, Maruff, Collie, & Darby, 2006) or to search for genetic associations (Haworth et al., 2007). The current study presents the first validation of a new self-administered web-based test battery, Memoro. The battery is designed to make large-scale assessments of memory and related cognitive functions in older adults and seniors using tests similar to validated neuropsychological tests frequently used by researchers and clinicians.

Computerized test batteries offer many potential advantages compared to traditional "pencil and paper" tests (American Psychological Association, APA, 1986). Computerized batteries can cover several cognitive domains with improved standardization of stimuli presentation and response collection (Wild, Howieson, Webbe, Seelye, & Kaye, 2008). Cognitive testing with computerized batteries has been rated as less difficult and less distressing (Collerton et al., 2007) and may increase efficiency and reduce costs of administration (Bauer et al., 2012). Although there has been a significant evolution in computerized batteries, many are still not fully compatible with the existing infrastructure of clinics and research institutions. Some batteries do, for example, require certain proprietary or special hardware (Silverstein et al., 2007; Wild et al., 2008), which is expensive and requires technical skill to install and operate (Collie, Darby, & Maruff, 2001).

Web-based neuropsychological test systems provide several unique possibilities both to research and to the clinic. Most computerized neuropsychological tests and traditional "pencil and paper" tests can be adapted to a web-based format and accessed through a single application, the web browser (Silverstein et al., 2007). To the clinician and researcher this means there is no need for buying new or special equipment to use these computerized tests. Patients and research participants can complete the tests at home using a familiar user interface without the need to download, install, and uninstall any software on their private computers, tablets, or smart phones. Several web-based neuropsychological test batteries are used in assessment and screening of cognitive functions and decline in adults and seniors (Zygouris & Tsolaki, 2014). Generally these batteries require a test administrator—for example, CNS Vital Signs (Gualtieri & Johnson, 2006) and Cognitive Stability Index (CSI; Erlanger et al., 2002). Still, some batteries are completely self-administered—for example, Cognitive Function Test (CFT; Trustram Eve &

de Jager, 2014), Computerized Self Test (COGselftest; Dougherty et al., 2010), and CogState (Hammers et al., 2012; Maruff et al., 2009). Although being completely self-administered, several web-based test systems seem to lack features making them optimal for organized large-scale assessment, such as an administrative user interface for designing, deploying, and monitoring assessments. These features are, however, implemented in Memoro.

Administering and interpreting computerized and web-based testing are not without challenges. For example, computer familiarity may influence validity of test results as it has been found to be associated with test performance (Iverson et al., 2009). There are indications that computer familiarity is associated with better cognitive test performance, regardless of whether the test is administered on a computer or in a traditional pencil-and-paper fashion (Fazeli, Ross, Vance, & Ball, 2013). Computer familiarity may vary with age and is thus of particular concern with older participants that may have less experience using computers. However, seniors appear to be able to follow instructions and self-administered computerized tests, making such tests feasible to assess cognitive function in all age groups (Collerton et al., 2007; Darby et al., 2014). Furthermore, technical aspects of the computer equipment used to administer the tests may influence performance and possibly validity. Depending on how tests are designed, even minor changes in hardware or user interface may have significant effects (Schlegel & Gilliland, 2007; Wild et al., 2008). It is important to reassess validity when adapting traditional or computerized neuropsychological tests for web-based administration since an adaptation can be seen as a new test (Bauer et al., 2012).

The primary aim of this study was to assess the validity of Memoro by comparing participants' performance on Memoro to that of a battery of traditional pencil-and-paper test analogues and to investigate how age and computer familiarity may influence participants' performance. We expected the participants to have consistent performances on the two batteries, with differing raw scores due to differences in test administration and response collection. We predicted a negative relationship between age and performance and a positive relationship between computer familiarity and performance. We also wanted to determine the feasibility of cognitive testing with Memoro in older and senior adults by evaluating the extent of missing data and participants' feedback on the test experience.

METHOD

Materials

The Memoro neuropsychological test battery

Memoro is a self-administered web-based neuropsychological test battery. Memoro is meant to measure memory and related cognitive functions in large cohorts using neuropsychological tests that are familiar to researchers and clinicians and at the same time provides flexibility of including new tests based on current research.

In order to protect privacy, no direct personally identifiable information is required to use Memoro. Memoro employs role-based privileges (e.g., administrator, researcher, assistant), ensuring that users only make authorized actions, which further secures the confidentiality of the test results. Through a web-based administrative user interface research administrators can create a project, generate a set of anonymous username/password combinations, monitor progress, and extract data. On the Memoro server all data are saved to a project-specific database securing separation of data between projects. All data traffic between the users and the server is encrypted.

To minimize the effects of variations in software/hardware configurations, we have designed the Memoro tests using cross-browser supported HTML and JavaScript. In the current battery no tests are dependent on precise millisecond precision. It should be noted though, that it is possible to measure small reaction time effects of ~20 ms on web-based platforms (Crump, McDonnell, & Gureckis, 2013). The Memoro tests are designed to be resistant to low bandwidth by preloading stimuli prior to presentation.

A concern to all self-administered testing is whether the participant completes the tests in a valid manner. Memoro employs context measures to register computer-, environment-, and participant-related threats to validity. Performance validity testing will flag exceptionally poor or good performances or unexpected differences in performance among tests. Normative data are being collected in order to incorporate this feature.

Participants

Sixty-five individuals agreed to participate in this study. Participants were recruited from health care, educational, and governmental organizations and through public poster boards in Trondheim County, Norway. To participate, individuals had

to sign up through a web-based registration page or call our project phone. The volunteers gave information on sex, age, and education, and answered questions regarding inclusion/exclusion criteria. The inclusion criterion was age ≥ 50 years, and exclusion criteria were previous or current neurological disease and participation in the HUNT population study (Krokstad et al., 2013) as this battery was intended to be administered in the HUNT study population. The volunteers were also asked whether they had psychiatric or medical conditions that may influence their performance. No participants had impairments in vision, hearing, or motor function. Participants were divided into two groups, one in which the Memoro tests were administered first and one in which the traditional tests were administered first in order to enable assessment with minimal of carry-over effects—that is, effect of order of testing. Assignment to groups was done quasirandomly striving to achieve equal levels of the demographic variables in the two groups. The study was evaluated by the Regional Committee for Medical and Health Research Ethics and approved by the Data Protection Official (Personvernombudet). Participants received a monetary reward of 50 USD after completing the last session.

Test administration traditional versus Memoro battery

Participants were tested in one of two identical rooms at the MR Center, St. Olavs University Hospital, Trondheim, Norway. Each room was set up with all the test materials for the traditional tests and a desktop computer with a 17" screen (resolution 1024 × 768 pixels) running Microsoft Windows XP and Mozilla Firefox (Version 15) for the Memoro tests. A standard 102-key keyboard, a two-button mouse, and speakers were attached to the computer.

Specific procedure for the traditional battery

Participants were given a short introduction to neuropsychological testing and how the administrator would behave during the testing session. Each participant was informed that the different tests were designed so that few or nobody would get everything correct. Some tests might appear easy and progress to become more difficult, while others might be perceived as difficult from the beginning.

Specific procedure for the Memoro battery

Each participant was logged into the Memoro system using a unique username/password combination. A research assistant gave a short instruction on how to operate the computer keyboard and mouse and made sure the sound volume was adjusted to the participants' preference. Participants were informed of the varying difficulty of the tests, in the same manner as for the traditional testing, and were given the opportunity to ask any questions before the research assistant left the room. Each Memoro test contained both verbal and written instructions. Instructions included information about how stimuli would be presented and how the participant was to use the keyboard or the mouse in order to respond.

Context measures

As a compulsory part of the Memoro test battery each participant started off his or her test session by completing a questionnaire about what kind of computer they were using, the physical location of that computer, and the noise level in the room. The participant was asked about sleep quantity and quality for the previous night and was asked to indicate how alert they felt. Lastly, a comment field was available for the participants to leave other information they might find relevant for their performance. These measures are included to determine whether the tests are completed in an acceptable context.

Neuropsychological measures

Verbal memory

Memoro: Verbal Memory Test. The participants were instructed to pay attention to the playback of audio recordings of word lists. There was a 2.5-s pause between each recorded word. The participants were instructed to use the keyboard to type in the words they recalled into text boxes on the screen. Participants first completed four learning/recall trials with a target list of 16 words having similar frequency of occurrence in the Norwegian language chosen from four different semantic categories. This was followed by one distraction learning/recall trial with a nontarget list of equal length and structure. After the distraction trial, participants completed an immediate recall trial with the original target list. Delayed free recall of the target list was performed after ~20 minutes of completing other nonverbal tests. Performance was scored as the number of correctly recalled words on the

immediate and the delayed free recall trials. There was no time limit on the recall parts of this test.

Traditional test: The California Verbal Learning Test Version 2 (CVLT-II) (Delis, Kramer, Kaplan, & Ober, 2000). The CVLT-II was chosen as a pen-and-paper analogue to the Verbal Memory Test. The test was administered and scored according to the manual.

Spatial memory

Memoro test: Objects in Grid. Participants were instructed to remember the location of 18 colored line drawings of various objects presented in a 6 × 6 grid. After a 90-s encoding phase the objects were automatically moved into two rows below the grid. The participants were instructed to use the mouse to drag and drop each object back to its correct position in the grid (immediate recall), and if they were uncertain about an object's location they were instructed to make a guess. After completing other nonspatial tests for ~15 min, participants were again presented with the empty grid and the objects in two rows below and were asked to drag and drop each object back to its correct position (delayed recall). This part was included first half-way through the study due to unforeseen delays in development. Performance was scored as the number of correctly placed objects separately for immediate and delayed recall. There was no time limit on the recall parts of this test.

Traditional test: Objects in Grid. The Objects in Grid Test is an adaptation of the Location Learning Test (LLT; Bucks & Willison, 1997). As a physical analogue we used a laminated paper variant of the 18 colored line drawings from the web-based test. The participants received the same instruction as that on the Memoro version, and the test administrator removed and shuffled the objects before immediate and delayed recall. Performance was scored as in the Memoro version.

Working memory

Memoro test: Digit Span Backwards. Participants were instructed to memorize a series of digits presented consecutively for 2 s each on the screen. After all digits in a trial had been presented, the participants were told to type the digits into a textbox in backwards order using the keyboard, and when finished click a button on the screen to continue to the next trial. The test consisted of 18 trials where the number of digits to remember increased by one for every second trial starting with two digits in the

first trial and ending with ten digits for the last trial. The test ended if a participant made three consecutive erroneous responses. The participants were not informed about this criterion. Performance was scored as number of correct responses.

Traditional test: Digit Span Backwards. The Digit Span Backwards subtest of the Wechsler Memory Scale 3rd edition (WMS-III; Wechsler, 1997) was administered according to the manual.

Memoro test: Letter-Number Sequencing. Participants were instructed to memorize a series of letters and digits presented consecutively on the screen for 2 s each. After all letters and digits in a trial had been presented, the participants were asked to organize first numbers in an ascending order and then the letters in alphabetical order and type the response in a textbox. The test consisted of 14 trials where the number of symbols to remember increased by one for every second trial starting with two symbols for the first trial and ending with eight symbols for the last trial. The test ended if a participant made three consecutive erroneous responses. The participants were not informed about this criterion. Performance was scored as number of correct responses.

Traditional test: Letter-Number Sequencing. The Letter-Number Sequencing subtest of the WMS-III (Wechsler, 1997) was administered according to the manual.

Processing speed

Memoro: Coding. Participants were presented with a matrix connecting geometrical symbols with single digit numbers (1-9) on the top of the screen. Below, the participants could see rows of symbols and empty cells below each symbol, with the first cell highlighted. The participants were asked to indicate which number was associated with the symbol above the highlighted cell and to press the corresponding numeric button on the keyboard. Correct responses resulted in a green color to the cell, while errors were marked with a red color. It was not possible to go back and correct erroneous responses. The participants were given 90 s to connect as many geometrical symbols and numbers as possible. The performance was scored as number of correct responses minus the number of erroneous responses.

Traditional test: Symbol-Digit Modalities Test (SDMT). The test was administered according to the manual (Smith, 1982).

Three additional tests (Card Sorting, Tower of London, and Word Pairs) were included in the battery but excluded from the current study due to problems with the tests design and stop criteria causing flooring effects and premature termination. The test sequence was (Memoro/traditional battery): Verbal Memory Test/CVLT-II, Coding/SDMT, Digit Span Backwards, Letter-Number Sequencing, Objects in Grid, Verbal Memory Test (late recall)/CVLT-II (late recall), Word Pairs, Tower test, Objects in Grid (late recall), Card Sorting, Word Pairs (late recall).

Computer familiarity

Within two weeks after completing both test packages the participants completed the Memoro Short Computer Questionnaire (MSCQ) by telephonic interview. The questionnaire contained six questions. Three assessing computer usage; “*Where have you used a computer during the last 6 months?*,” alternatives: “Home (2), Work (2), Other (1),” “*What activities have you done on a computer during the last 6 months?*,” alternatives: “Paying bills (1), E-mail (1), Browsing (1), Office (1), Multimedia (1),” “*How often to you use a computer?*,” alternatives: “Daily (5), Several times a week (4), Once a week (3), More than once a month (2), Less than once a month (1).” Three assessing computer skill; “*How comfortable are you in using a computer mouse?*,” alternatives: Very uncomfortable (1), Uncomfortable (2), Neither nor (3), Comfortable (4), Very comfortable (5), “*How comfortable are you in using a computer keyboard?*,” same alternatives as previous question, “*How comfortable are you at using computers on a scale from 1 to 10 with 1 being ‘only problems’ and 10 ‘no problems at all’?*.” Each participant got a computer use score, a computer skill score, and a combined total score used as a measure of computer familiarity. The total score was used for statistical analysis in this study.

Participant feedback

After completing the Memoro battery, participants completed a qualitative structured interview; “*How did you perceive the instructions?*,” response coded as negative, neutral, or positive, “*How did you find completing the tests in Memoro?*,” “*How did you find responding by drag-and-drop?*,” “*How did you find responding by typing?*,” “*How did you find navigating in and between tests?*,” responses coded as either “indicated difficulties/problems” or “no difficulties/problems” for each question. After the

TABLE 1
Participant feedback interview results

Question	Positive N (%)	Negative N (%)	Neutral N (%)
How did you perceive the Memoro instructions?	10 (18.5)	3 (5.6)	41 (75.9)
	Prefer Memoro	Prefer traditional	No preference
How did you find completing Memoro compared to the traditional tests?	24 (44.4)	13 (24.1)	17 (31.5)
Perceived Memoro battery as less difficult	11 (45.8)		
More in control of the test situation	7 (29.2)		
Not a person waiting for or evaluating responses	6 (25)		
Perceived traditional battery as less difficult		5 (38.4)	
Less exhausting with traditional tests		4 (30.8)	
Better to relate to a person		4 (30.8)	
	No problems	Problems	
How did you find completing the tests in Memoro?	50 (92.6)	4 (7.4)	
How did you find responding by drag-and-drop?	47 (87)	7 (13)	
How did you find responding by typing?	53 (98.1)	1 (1.9)	
How did you find navigating in and between tests?	54 (100)		

Note. N = number of subjects; % = percentage of all subjects answering.

second session when participants had completed both batteries they were asked “*How did you find completing Memoro compared to the traditional tests?*,” with responses coded as “Preferred traditional tests,” “Perceived equal,” “Preferred Memoro.” Reasons for preference were coded as shown in Table 1.

HUNT dataset used for extended analysis of the Memoro battery

To investigate the relationship between age, education, and computer familiarity and Memoro performance, and run factor analyses in an appropriately sized dataset we included data from the HUNT study. The HUNT dataset is from a subgroup ($N = 1006$) of individuals who participated in the public health surveys in Nord-Trøndelag county (HUNT 1, years 1985 to 1987; HUNT 2, years 1995 to 1997; HUNT 3, years 2006 to 2008; HUNT MRI, years 2007 to 2009; Honningsvåg, Linde, Håberg, Stovner, & Hagen, 2012). From this group of individuals a subset was recruited for trying out web-based cognitive testing in general population studies. The testing was performed in 2013 to 2014. Participants received an invitation including login information by mail. Participants could perform the tests at home, or book a session at the HUNT research center where we had set up computers for them to use. Tests included and used in this study were: Verbal Memory Test, Objects in Grid, Letter–Number Sequencing, Processing Speed, Verbal Memory

Test (late recall), and Objects in Grid (late recall). The processing speed task was different than in the Memoro versus traditional test battery validation sample. Participants completed six trials of 30 s and were instructed to judge as fast as possible without making mistakes whether pairs of numbers or geometrical shapes were identical or different by hitting the “F” or “L” key on the keyboard. Performance was scored as number of correct responses minus number of erroneous responses. The remaining tests were identical in the Memoro versus traditional test battery validation sample and the HUNT sample. In total, 228 persons from HUNT completed all Memoro tests included. Of these, 10 participants indicated having problems completing the tests; hence 218 participants were included in the final analyses.

Statistical analysis

From the Memoro versus traditional battery validation sample and the HUNT sample raw scores were extracted from the Memoro database into a SPSS data file. The raw scores from the traditional tests were punched into the same file. Data were labeled as missing if no response had been registered by the system or if the subject clearly had misunderstood the instruction. Differences in demographics in the Memoro first versus traditional first groups were investigated by independent-samples t tests except for education where the Mann–Whitney U -test was used. Raw score differences were analyzed using paired-sample t

tests. Carry-over (order) effects were investigated using independent-samples *t* tests. Effect sizes were calculated and expressed as Cohen's *d* using pooled variance. Concurrent validity was assessed by correlating the Memoro and traditional test scores using Pearson correlation. Factor structures for Memoro and traditional test battery were investigated using exploratory factor analysis with unweighted least squares extraction. Oblique rotation (oblimin) was used as we expected the different cognitive constructs to be interrelated. Kaiser's criterion ($\lambda > 1$) and scree plots were used to determine number of factors (Field, 2009). Associations between age, computer familiarity, and test performance were investigated using Pearson correlation. Association between participant feedback and test performance was investigated using independent-sample *t* tests. In the HUNT sample, factor structure for the Memoro tests was investigated using the same methods as those above. Differences between the HUNT sample and the study sample in age and computer familiarity were investigated using independent-samples *t* tests while Mann–Whitney was used for education. The relationship between age, education, computer familiarity, and test performance in this larger sample was investigated using Pearson and Spearman (education) correlation methods. Analysis of variance (ANOVA) with Bonferroni post hoc tests were used to investigate possible differences in performance among participants in the HUNT sample taking tests at different locations.

Statistical analyses were performed with IBM SPSS Statistics for Windows (Version 20) and R Version 3.1.1. Probabilities of $p < .05$ (two-tailed) were considered statistically significant, after application of the Bonferroni–Holm (Holm, 1979) method to correct for multiple comparisons.

RESULTS

Participants in Memoro versus traditional battery validation sample

Initially 65 (39 female) individuals agreed to participate, but two females withdrew their consent, and one male did not show up for the second session. Hence, 62 participants (37 females, mean age = 60.7 ± 7.1 years, median education = 4, MSCQ = 29.6 ± 4.6) were included in the statistical analyses. The Memoro first and traditional first groups did not differ significantly on measures of age, education, gender distribution, or computer familiarity measured by MSCQ (Table 2). No significant carry-over (order) effects in test performance were observed between the groups ($t = -0.294$ to 1.866 , $df = 30$ to 60 , all $p > .05$). This was expected as we used a counterbalanced design. The median time between the two sessions was 15 (13–27) days. None of the participants reported issues in the context measures (e.g., high noise level or not feeling alert) that might threaten test validity.

Participants in the HUNT sample

A total of 218 participants from the HUNT population study were included in the analyses (Table 2). The HUNT sample had significant lower computer familiarity, $t(273) = 6.665$, $p < .001$, $d = 1$, 95% CI $[-6.75, -3.67]$, higher age, $t(273) = 3.547$, $p < .001$, $d = 0.45$, 95% CI $[1.22, 4.09]$, and lower median education, $u = 3877.5$, $p < .001$, $r = .27$, than the validation sample. A total of 76.6% completed the tests at home without other people present in the same room, 13.1% completed the tests at the HUNT research center, 7.5% completed the tests at home with other people present in the same room, and 2.8% completed the tests at work without other

TABLE 2
Participant characteristics

Measure	Memoro first			Traditional first			Memoro (HUNT)		
	<i>n</i>	Mean (<i>SD</i>)	Range	<i>n</i>	Mean (<i>SD</i>)	Range	<i>n</i>	Mean (<i>SD</i>)	Range
Age		60.8 (6.2)	52–77		60.6 (8.1)	50–80		63.38 (4.31)	56–72
Edu. ^a		4 (3–5) ^b	1–5		4 (3–5) ^b	2–5		3 (2–4) ^b	1–5
CF		29.7 (3.2)	21–34		29.5 (5.8)	9–35		24.42 (5.4)	5–33
Sex (M/F)	13/19			12/18			111/107		

Note. Edu. = education; CF = computer familiarity; M = male; F = female.

^aEducation values represent completed: 1, secondary school; 2, high school; 3, ≤ 3 years at college/university; 4, ≤ 5 years at college/university; 5, ≥ 5 years at college/university.

^bCentral tendency is presented as median with interquartile range.

TABLE 3
Correlations between and within the Memoro and traditional batteries

Measure	Comparison	VM IR	VM DR	SM IR	SM DR	WM DSB	WM LNS	PS
VM IR	M,T	.51**	.48**	.31	.33	.07	.10	.22
	M,M	1	.89**	.41*	.37	.36	.16	.31
	T,T	1	.87**	.55**	.56**	.27	.28	.30
VM DR	M,T	.51**	.52**	.29	.17	.15	.05	.25
	M,M	.89**	1	.42*	.37	.34	.14	.33
	T,T	.87**	1	.57**	.55**	.21	.21	.27
SM IR	M,T	.48**	.49**	.63**	.58**	.27	.18	.24
	M,M	.41*	.42*	1	.87**	.22	.20	.52**
	T,T	.55**	.57**	1	.90**	.13	.08	.29
SM DR	M,T	.65**	.66**	.70**	.63**	.27	.19	.46*
	M,M	.37	.37	.87**	1	.39	.28	.42
	T,T	.56**	.55**	.90**	1	.04	.27	.21
WM DSB	M,T	.34	.30	.20	.24	.49**	.56**	.40*
	M,M	.36	.34	.22	.39	1	.55**	.41*
	T,T	.27	.21	.13	.04	1	.63**	.37**
WM LNS	M,T	.43*	.40*	.19	.31	.48**	.54**	.43*
	M,M	.16	.14	.20	.28	.55**	1	.40*
	T,T	.28	.21	.08	.27	.63**	1	.31
PS	M,T	.24	.28	.22	.17	.42*	.44*	.49**
	M,M	.31	.33	.52**	.42	.41*	.40*	1
	T,T	.30	.27	.29	.21	.37*	.31	1

Note. Numbers in bold represent concurrent validity comparisons. Comparison column first letter represents rows, second letter represents columns (e.g., M,T = Memoro in rows, traditional battery in columns). T = traditional battery; M = Memoro; VM = verbal memory; SM = spatial memory; IR = immediate recall; DR = delayed recall; WM = working memory; DSB = digit span backwards; LNS = letter-number sequencing; PS = processing speed. Coefficients are Pearson *r*.
*Adjusted *p* < .05. **Adjusted *p* < .01.

people present in the same room. ANOVA Bonferroni post hoc tests showed no statistical differences in performance on any of the tests between the different test locations.

Concurrent validity

Moderate correlations were observed between the Memoro and traditional test performance measures. Correlation coefficients ranged from .63 for Objects in Grid to .49 for Digit Span Backwards and Processing Speed when comparing the

Memoro tests to their traditional test analogues (Table 3).

Factor structures in Memoro versus traditional test validation sample

Exploratory factor analysis was performed on the raw scores of the Memoro tests and the traditional tests separately (Table 4). Two factors were extracted among the Memoro tests, *Spatial Memory* ($\lambda = 3.250$) and *Verbal Memory* ($\lambda = 1.309$). Similarly, two factors were extracted among the traditional

TABLE 4
Exploratory factor analysis with Oblimin rotation, factor structure (correlations)

Measure	Memoro		Traditional		Memoro (HUNT)	
	VM	SM	VM	SM	VM	SM
VM IR	.983	.456	.909	.602	.946	.143
VM DR	.924	.409	.896	.597	.948	.170
SM IR	.390	.861	.616	.928	.139	.853
SM DR	.459	.997	.573	.935	.165	.945
WM DSB	.386	.375	.403		NA	NA
WM LNS	.242	.269	.375	.236	.173	.246
PS		.370	.345	.234		.450

Note. CF = Computer familiarity; VM = verbal memory; SM = spatial memory; IR = immediate recall; DR = delayed recall; WM = working memory; DSB = digit span backwards; LNS = letter-number sequencing; PS = processing speed. NA = not applicable as this test was not included in the HUNT sample.

tests, *Spatial Memory* ($\lambda = 3.416$) and *Verbal Memory* ($\lambda = 1.514$). The correlation between the factors was .46 for the Memoro factors and .62 for the traditional battery factors. The sampling adequacy for the set of variables was acceptable (Kaiser–Meyer–Olkin; Memoro = .613, traditional tests = .628), yet the Tucker Lewis index of the obtained factor solution suggested a poor model fit (Memoro = .756, traditional tests = .568). This indicates that the results from this factor analysis must be interpreted with caution.

Factor structures in the HUNT sample

The factor analysis in the larger HUNT dataset gave two factors; *Spatial Memory* ($\lambda = 2.449$) and *Verbal Memory* ($\lambda = 1.724$; Table 4), with an inter-correlation of .17. The sampling adequacy for this analysis was low (Kaiser–Meyer–Olkin: .570), and the model fit was marginally good (Tucker Lewis index: .941). The impact of a larger sample on separating spatial and verbal memory abilities based on traditional testing could unfortunately not be performed in the HUNT sample as it lacked pen and paper testing.

Performance characteristics

As expected, the scores for each Memoro test and its traditional analogue differed significantly, with effect sizes ranging from small to very large in the Memoro versus traditional test validation sample (Table 5). The largest difference was observed between the processing speed tests ($d = -2.54$). Performance was not systematically better for any

battery across the different tests. This is clearly demonstrated by the finding that participants performed better on the verbal memory test and worse on the spatial memory test in the Memoro battery than in the traditional battery.

Computer familiarity, age, education, and test performance

In the Memoro versus traditional test validation sample we found computer familiarity to be moderately positively correlated with performance on all Memoro tests and the processing speed test of the traditional battery. Age was found to be negatively correlated to some tests on both the Memoro and traditional batteries. Computer familiarity showed a negative correlation to age, and there were no significant correlations between computer familiarity and test performance after adjusting for age (Table 6). Education level was not correlated with either age or computer familiarity and was not correlated to any of the performance measures. In the HUNT sample, computer familiarity correlated only with the processing speed task. Both age and education correlated with computer familiarity in this sample. Please note the large discrepancy in the correlation coefficients for age between the two samples in this study. While age only showed a negative correlation to processing speed, education correlated significantly with all tests except the verbal memory tests. After adjusting for age and education, there were no significant correlations between computer familiarity and test performance.

TABLE 5
Performance on the Memoro and traditional test battery

Measure	Memoro		Traditional		<i>d</i>	Memoro (HUNT)	
	<i>N</i> (%)	Mean (<i>SD</i>)	<i>N</i> (%)	Mean (<i>SD</i>)		<i>N</i> (%)	Mean (<i>SD</i>)
VM IR	60 (96.8)	13.00 (2.78)	61 (98.4)	11.18 (3.53)**	0.57	218 (100)	12.76 (2.8)
VM DR	62 (100)	13.23 (2.64)	61 (98.4)	11.51 (3.59)**	0.55	218 (100)	12.99 (2.68)
SM IR	60 (96.8)	10.77 (4.31)	62 (100)	12.08 (4.15)*	-0.31	218 (100)	9.06 (3.85)
SM DR ^a	32 (94.1)	8.69 (5.38)	38 (100)	10.55 (4.57)*	-0.37	218 (100)	7.76 (3.9)
WM DSB	61 (98.4)	8.68 (2.92)	62 (100)	6.90 (1.89)**	0.72	NA	NA
WM LNS	62 (100)	8.29 (2.57)	62 (100)	9.23 (3.2)*	-0.32	218 (100)	7.90 (2.83)
PS	58 (93.5)	29.31 (9.05)	62 (100)	60.01 (14.5)**	-2.54	218 (100)	46.76 (12.65)

Note. VM = verbal memory; SM = spatial memory; IR = immediate recall; DR = delayed recall; WM = working memory; DSB = digit span backwards; LNS = letter–number sequencing; PS = processing speed; *N* = number of responses; % = percentage of valid responses; NA = not applicable as test was not included in the HUNT sample.

^aSpatial memory: Objects in Grid late recall was administered to 34 subjects on Memoro and 38 subjects on the traditional battery as this part was included halfway through the study.

*Paired-samples *t* test (Memoro–traditional) adjusted $p < .05$. **Paired-samples *t* test (Memoro–traditional) adjusted $p < .01$.

TABLE 6
Computer familiarity, age, education, and test result correlations

Measure	Memoro				Traditional				Memoro (HUNT)			
	Age	Edu. ^a	CF	CF ^b	Age	Edu. ^a	CF	CF ^a	Age	Edu. ^a	CF	CF ^c
Age	1	-.15	-.62**	0	1	-.15	-.62**	0	1	.13	-.17*	0
Edu.	-.15	1	.18	-.02	-.15	1	.18	-.02	.13	1	.30**	0
VM IR	-.29	.21	.34*	.20	-.10	.07	.24	.24	-.09	.09	-.01	-.07
VM DR	-.30*	.14	.32*	.16	-.07	.02	.29	.26	-.07	.02	-.02	-.04
SM IR	-.29	.08	.40**	.31	.07	.04	.04	.16	-.13	.20*	.22	.15
SM DR	-.22	.16	.43*	.35	.15	.19	.12	.27	-.13	.22**	.22	.14
WM DSB	-.41*	.24	.37*	.15	-.34*	.14	.30	.14	NA	NA	NA	NA
WM LNS	-.26	.16	.36*	.28	-.38*	.18	.24	.02	-.09	.17*	.18	.07
PS	-.63**	.16	.53**	.18	-.43**	.13	.47**	.26	-.26**	.23**	.31**	.19

Note. Edu. = education; CF = computer familiarity; VM = verbal memory; SM = spatial memory; IR = immediate recall; DR = delayed recall; WM = working memory; DSB = digit span backwards; LNS = letter-number sequencing; PS = processing speed. NA = not applicable as test was not included in the HUNT sample. Coefficients are Pearson r .

^aSpearman correlation coefficients.

^bAdjusted for age.

^cAdjusted for age and education.

*Adjusted $p < .05$. **Adjusted $p < .01$.

Missing data and participant feedback

For the traditional tests, missing data were only observed on verbal memory (98.4% completion). The Memoro tests had more missing data. Verbal memory immediate recall and spatial memory immediate recall had 96.8% completion, spatial memory delayed recall had 94.1% completion, and the processing speed test had 93.5% completion. Three tests were excluded from this study because of too strict stop criteria (Memoro tower test: 25% missing data; Memoro card sorting test: 48% missing data) and too few encoding trials (Memoro word pairs task), resulting in a flooring effect. The tests had been piloted in younger subjects where no flooring was observed. Premature termination of the Tower test could influence the duration between immediate and delayed recall of the spatial memory task and consequently performance on delayed recall. However, using data from the HUNT sample for comparison, we found no statistically significant difference in the mean difference between immediate and delayed recall in the two datasets, $t(264) = 1.548$, $p = .123$, $d = 0.28$, 95% CI [-0.172, 1.435].

Feedback from 54 participants from the Memoro versus traditional validation sample was analyzed (Table 1). Six participants did not complete the interview, and two had only partial completion due to time constraints. Independent-samples t tests between the participants that preferred the traditional battery ($n = 13$) and those that preferred Memoro ($n = 24$) indicated no significant difference on any of the traditional tests, Memoro tests, age, or computer familiarity.

DISCUSSION

In this study we have presented a new self-administered web-based neuropsychological test battery for large-scale assessment of memory and related cognitive functions in adults and seniors.

We found moderate correlations between the Memoro tests and the traditional tests presumed to measure similar constructs demonstrating good concurrent validity. A review of the literature of concurrent validity of computerized neuropsychological tests reported median correlations of .40 to .46 for memory, .28 to .40 for psychomotor speed, .41 to .48 for executive function, .24 to .56 for attention, and .50 to .60 for reaction time (Gualtieri & Johnson, 2006). Memoro came out in the upper range when compared to more similar web-based tests. For example, for memory we found correlations in the .5 to .6 range. This is in the high end of previous reports. The CSI memory factor had a .52 correlation to the Buschke Selective Reminding Test (Erlanger et al., 2002), CFT episodic memory recognition had a .39 correlation to Total Doors and People, and CFT episodic memory recall had a .43 correlation to People recall (Trustram Eve & de Jager, 2014). One study found the CogState One-Back and Learn tasks to correlate .54 and .83 to the Brief Visual Memory Test (Maruff et al., 2009). Another study using the Visual Reproduction test of the Wechsler's Memory Scale (WMS) found correlations in the .45 to .50 range for the CogState Detect, Identify, and One-Back tasks but not the Learning task (Hammers et al., 2012). Direct comparisons of our results to that of other studies

should be interpreted with caution because other batteries include other tests and have other data considerations affecting their analyses. However, our findings do indicate that Memoro performs as well as previously validated and well-used web-based test batteries aimed at assessing cognitive function in adults and seniors.

The factor analyses suggested that both the Memoro and traditional battery could be separated into two factors: Verbal Memory and Spatial Memory. However, there was a relatively large overlap between the factors, and the solution did not fit the data well. In the more appropriately sized HUNT sample we obtained the same verbal and spatial memory factors in Memoro, with less overlap between the factors and with a better model fit. This finding supports that verbal and spatial memory may to some degree be separable. However, we acknowledge that both analyses have methodological weaknesses and that additional work in larger samples is needed to validate the obtained factor structure.

The differences in raw scores combined with consistent performance observed between Memoro and traditional test batteries were expected as any adaptation of a traditional test to a computerized version results in a new test (Bauer et al., 2012). The Memoro tests were not intended to be identical to the tests in the traditional battery, but to measure the same cognitive constructs. Differences in test administration, response collection, stop criteria, and perceived difficulty can account for differences in raw scores between traditional and web-based tests. The largest difference in raw scores was observed between the processing speed tasks. The relatively lower score on the processing speed task in Memoro most likely stemmed from the participants having to switch between looking at the keyboard, the symbol-digit key, and the current stimulus to be attended on the screen, resulting in slower and in consequence fewer responses. Similar speed differences between traditional testing and computerized testing due to differences in response method have been shown previously. For example, the mean raw score on a working memory reaction time task decreased by 45% when the response mode was changed from pointing at a touch screen to using the computer mouse (Silverstein et al., 2007).

Notably, the verbal memory performance was higher when measured with Memoro than with the traditional test, whereas the opposite was found for spatial memory. In line with the above arguments, this flip-flop may be explained by differing response methods. On the verbal memory test, participants responded by typing the words

in Memoro, compared to saying them out loud in the traditional battery. The Memoro response method might strengthen encoding in that the participants see the words they typed adding both a second modality (vision) and possibly some working memory relief. Interestingly, the higher score on the Memoro verbal memory test was present even though there were only four learning/recall trials and not five as in CVLT-II. We tried a fifth learning/recall trial during piloting, but this contributed to pronounced ceiling effects. On the spatial memory task, Memoro required participants to drag and drop objects with the mouse, which may have been more difficult than placing cards on a board as in the traditional test. The perceptual differences of having stimuli presented on a screen versus a table top may also have influenced the scores. Our findings show that although the Memoro tests appear to be similar to their traditional analogues, differences in how the tests are perceived and responded to cause shifts in raw scores. This could cause an interpretative problem. We could question whether we measure the same construct with a test that is perceived as more easy/difficult than its traditional analogue, and whether it will impact, for example, the test's sensitivity and specificity. The fact that the two test batteries yielded similar factor structures can be taken to suggest that they measure the same cognitive constructs, but the results of these factor analyses must await further confirmation. Nonetheless, the flip-flop highlights the fact that raw scores from new tests may not be interchangeable with that of its traditional analogues. Test-specific and administration-specific norms must be developed and psychometric properties investigated before a new test can be used for neuropsychological profiling or screening for clinical symptoms.

We found positive correlations between the measure of computer familiarity and performance on both the Memoro and the traditional battery. Previous work has shown computer attitudes (Weber, Fritze, Schneider, Kühner, & Maurer, 2002), computer-related anxiety (Brownlyke et al., 2002), and self-reported computer familiarity (Iverson et al., 2009) to influence both computerized and traditional (Fazeli et al., 2013) test performance. In the Memoro versus traditional validation sample there was a strong relationship between age and computer familiarity not observed in the larger HUNT sample. We believe the reason for this discrepancy to be the more narrow age range and lower education level in the HUNT sample. Education was found to have significant positive correlations to all except the verbal

memory test on the Memoro battery in the HUNT sample. We found no significant correlations between education and test performance on either the Memoro tests or the traditional tests in the Memoro versus traditional validation sample. This difference between the two samples can be explained by both the larger size and variability in education level in the HUNT sample. After adjusting for the possible confounding effects of age in the Memoro versus traditional validation sample, and age and education in the HUNT sample, we no longer observed any effects of computer familiarity on test performance. Taken together, these findings show that performance on both web-based and traditional tests may be influenced by the participants' self-reported computer familiarity, but these effects can be explained by confounding variables, such as age and education. Our findings also highlight that adjustment for computer familiarity should be done carefully as it might be related to other constructs of interest.

The majority of participants reported no problems understanding the instructions and completing the Memoro tests. This is in line with previous findings that older adults and seniors can successfully complete computerized and web-based tests (Collerton et al., 2007; Fredrickson et al., 2010). As some experienced difficulties with the drag-and-drop response method, it should be considered whether this response method may introduce more missing data and problems when used in a population of older individuals. Almost twice as many participants preferred the Memoro battery to the traditional battery. Perceiving the battery as less difficult was the main reason for preferring both the traditional as well as the Memoro battery. Those preferring Memoro appreciated the increased control they had in the test situation, and not being scrutinized by a test administrator. The opposite was observed for those preferring the traditional battery as they were more relaxed and appreciated the opportunity to relate to a person during testing. Comparing those who preferred Memoro to those preferring the traditional battery, we found no significant differences in test performance, age, or computer familiarity. This could suggest that personality characteristics rather than performance characteristics matter in preference of battery.

There are several limitations in this study that need to be addressed. First, we experienced unexpected technical problems with the test design and stop criteria on three tests not discovered during piloting. These tests were excluded from the analyses pending changes and consideration of future inclusion. As a result, the battery currently does

not include a specific measure of executive function. For the Memoro battery to assess more aspects of memory and related as well as other cognitive constructs, inclusion of measures of recognition, language skills, and visuoconstruction is needed. This initial version of the battery was, however, intended as a core package to which new tests can be incorporated in the future. In addition, some participants accidentally terminated the Memoro coding test. This was the only test where subjects got visual feedback on the screen about the accuracy of their responses, and four participants pressed the backspace key as a reflex to go back and correct a wrong response. This action terminated the test. Based on this experience, we have revised the test logic and programmatically disabled the backspace key to avoid this source of missing data in the future. The battery had been piloted in over 100 subjects aged <50 years without the problems of premature termination of tests experienced in this study.

Second, participants in this study were given some initial instructions by a research assistant in how to operate the computer keyboard, mouse, and adjustment of the sound volume before starting the tests. This was done as we did not expect all the participants to be familiar with our computer set-up. Such help will not be available if the tests are completed at home or some other offsite location. We have implemented specific audio instructions and on-screen animations to aid participants in such instances (e.g., in the HUNT sample used in this study). This could have influenced performance; however, in the HUNT sample we found no significant difference in performance among those participants who completed the tests at home, work, or at the HUNT research center where we had an assistant.

Third, the sample had a relatively high level of education. This may challenge generalization of our results to groups with lower levels of education. As Trondheim is a university city with 40.1% of the population having completed higher (tertiary) education (Statistics Norway, 2014), the high education level in the present sample is not unexpected. As we had a significantly lower education level in the HUNT sample this limitation does not apply to the results from this sample.

Fourth, the factor analyses in the Memoro versus traditional test sample was influenced by the low sample size and had several weaknesses that warrants caution in interpretation of results. The factor analysis in the HUNT sample had a larger sample size and showed a better fit, but we note that Kaiser–Meyer–Olkin around .6 is in the lower range of what is considered acceptable.

Our study shows that web-based neuropsychological test systems like Memoro can make significant contributions in cognitive assessments. Above all, they make it possible to perform efficient and standardized assessment in large cohorts. Good concurrent validity between the Memoro tests and the traditional tests, despite some technical problems leading to missing data, demonstrates that web-based testing with Memoro can be successfully implemented. Further work is necessary to ensure validity, reliability, and other psychometric properties of Memoro, both in lab settings and in more naturalistic home settings.

REFERENCES

- American Psychological Association. (1986). *Guidelines for computer-based tests and interpretations*. Washington, DC: American Psychological Association.
- 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. *The Clinical Neuropsychologist*, 26(2), 177–196. doi:10.1080/13854046.2012.663001
- Browndyke, J. N., Albert, A. L., Malone, W., Schatz, P., Paul, R. H., Cohen, R. A., ... Gouvier, W. D. (2002). Computer-related anxiety: Examining the impact of technology-specific affect on the performance of a computerized neuropsychological assessment measure. *Applied Neuropsychology*, 9(4), 210–218. doi:10.1207/S15324826AN0904_3
- Bucks, R. S., & Willis, J. R. (1997). Development and validation of the Location Learning Test (LLT): A test of visuo-spatial learning designed for use with older adults and in dementia. *The Clinical Neuropsychologist*, 11(3), 273–286. doi:10.1080/13854049708400456
- Collerton, J., Collerton, D., Arai, Y., Barrass, K., Eccles, M., Jagger, C., ... Kirkwood, T. (2007). A comparison of computerized and pencil-and-paper tasks in assessing cognitive function in community-dwelling older people in the Newcastle 85+ Pilot Study. *Journal of the American Geriatrics Society*, 55(10), 1630–1635. doi:10.1111/j.1532-5415.2007.01379.x
- Collie, A., Darby, D., & Maruff, P. (2001). Computerized cognitive assessment of athletes with sports related head injury. *British Journal of Sports Medicine*, 35(5), 297–302. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1724391&tool=pmcentrez&rendertype=abstract>
- Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PloS One*, 8(3), e57410. doi:10.1371/journal.pone.0057410
- Darby, D. G., Fredrickson, J., Pietrzak, R. H., Maruff, P., Woodward, M., & Brodtmann, A. (2014). Reliability and usability of an internet-based computerized cognitive testing battery in community-dwelling older people. *Computers in Human Behavior*, 30, 199–205. doi:10.1016/j.chb.2013.08.009
- Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *California Verbal Learning Test* Second Edition. *Adult Version. Manual*. San Antonio, TX: The Psychological Corporation.
- Dougherty, J. H., Cannon, R. L., Nicholas, C. R., Hall, L., Hare, F., Carr, E., ... Arunthamkun, J. (2010). The computerized self test (CST): An interactive, internet accessible cognitive screening test for dementia. *Journal of Alzheimer's Disease*, 20(1), 185–195. doi:10.3233/JAD-2010-1354
- Erlanger, D. M., Kaushik, T., Broshek, D., Freeman, J., Feldman, D., & Festa, J. (2002). Development and validation of a web-based screening tool for monitoring cognitive status. *The Journal of Head Trauma Rehabilitation*, 17(5), 458–476. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12802255>
- Falleti, M. G., Maruff, P., Collie, A., & Darby, D. G. (2006). Practice effects associated with the repeated assessment of cognitive function using the CogState battery at 10-minute, one week and one month test-retest intervals. *Journal of Clinical and Experimental Neuropsychology*, 28(7), 1095–1112. doi:10.1080/13803390500205718
- Fazeli, P. L., Ross, L. A., Vance, D. E., & Ball, K. (2013). The relationship between computer experience and computerized cognitive test performance among older adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 68(3), 337–346. doi:10.1093/geronb/gbs071
- Field, A. P. (2009). Exploratory factor analysis. In *Discovering statistics using SPSS* (3rd ed., pp. 627–685). London: SAGE.
- Fredrickson, J., Maruff, P., Woodward, M., Moore, L., Fredrickson, A., Sach, J., & Darby, D. (2010). Evaluation of the usability of a brief computerized cognitive screening test in older people for epidemiological studies. *Neuroepidemiology*, 34(2), 65–75. doi:10.1159/000264823
- 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. doi:10.1016/j.acn.2006.05.007
- Hammers, D., Spurgeon, E., Ryan, K., Persad, C., Barbas, N., Heidebrink, J., ... Giordani, B. (2012). Validity of a brief computerized cognitive screening test in dementia. *Journal of Geriatric Psychiatry and Neurology*, 25(2), 89–99. doi:10.1177/0891988712447894
- Haworth, C. M. A., Harlaar, N., Kovas, Y., Davis, O. S. P., Oliver, B. R., Hayiou-Thomas, M. E., ... Plomin, R. (2007). Internet cognitive testing of large samples needed in genetic research. *Twin Research and Human Genetics*, 10(4), 554–563.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65–70.
- Honningsvåg, L.-M., Linde, M., Håberg, A., Stovner, L. J., & Hagen, K. (2012). Does health differ between participants and non-participants in the MRI-HUNT study, a population based neuroimaging study? The Nord-Trøndelag health studies 1984–2009. *BMC Medical Imaging*, 12, 23. doi:10.1186/1471-2342-12-23
- 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. doi:10.1080/13803390802372125
- Krokstad, S., Langhammer, A., Hveem, K., Holmen, T. L., Midthjell, K., Stene, T. R., . . . Holmen, J. (2013). Cohort Profile: The HUNT Study, Norway. *International Journal of Epidemiology*, 42(4), 968–977. doi:10.1093/ije/dys095
- Maruff, P., Thomas, E., Cysique, L., Brew, B., Collie, A., Snyder, P., & Pietrzak, R. H. (2009). Validity of the CogState brief battery: Relationship to standardized tests and sensitivity to cognitive impairment in mild traumatic brain injury, schizophrenia, and AIDS dementia complex. *Archives of Clinical Neuropsychology*, 24(2), 165–178. doi:10.1093/arclin/acp010
- Schlegel, R. E., & Gilliland, K. (2007). Development and quality assurance of computer-based assessment batteries. *Archives of Clinical Neuropsychology*, 22 (Suppl. 1), S49–S61. doi:10.1016/j.acn.2006.10.005
- Silverstein, S. M., Berten, S., Olson, P., Paul, R., Willams, 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. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/18183911>
- Smith, A. (1982). *Symbol Digits Modalities Test*. Los Angeles, CA: Western Psychological Services.
- Statistics Norway. (2014). *Population's level of education, 1 October 2013*. Retrieved from <http://www.ssb.no/en/utdanning/statistikker/utniv/aar/2014-06-19?fane=tabell&sort=nummer&tabell=181264>
- Trustram Eve, C., & de Jager, C. A. (2014). Piloting and validation of a novel self-administered online cognitive screening tool in normal older persons: The Cognitive Function Test. *International Journal of Geriatric Psychiatry*, 29(2), 198–206. doi:10.1002/gps.3993
- Weber, B., Fritze, J., Schneider, B., Kühner, T., & Maurer, K. (2002). Bias in computerized neuropsychological assessment of depressive disorders caused by computer attitude. *Acta Psychiatrica Scandinavica*, 105(2), 126–130. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11939962>
- Wechsler, D. (1997). *Wechsler Memory Scale* (3rd ed.). San Antonio, TX: Psychological Corporation.
- Wild, K., Howieson, D., Webbe, F., Seelye, A., & Kaye, J. (2008). Status of computerized cognitive testing in aging: A systematic review. *Alzheimer's & Dementia*, 4(6), 428–437. doi:10.1016/j.jalz.2008.07.003
- Zygouris, S., & Tsolaki, M. (2014). Computerized cognitive testing for older adults: A review. *American Journal of Alzheimer's Disease and Other Dementias*. doi:10.1177/1533317514522852