Psychometric properties for a comprehensive cognitive test battery, traditional cognitive tests adapted for self-administration on a digital platform

Poster · August 2021

DOI: 10.13140/RG.2.2.33419.36647

4 authors, including:

Wobbie van den Hurk
Mindmore AB
6 PUBLICATIONS **34** CITATIONS
SEE PROFILE

Katarina Nägga
Lund University
111 PUBLICATIONS **3,700** CITATIONS
SEE PROFILE

Psychometric properties for a comprehensive cognitive test battery, traditional cognitive tests adapted for self-administration on a digital platform

AUTHORS & INSTIUTIONS

Wobbie van den Hurk[1,2], Frida Strömgren[1], Anders Gustavsson[3,4] and Katarina Nägga[2]

1 - Mindmore AB, Stockholm, Sweden
2 - Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden
3- Quantify Research, Stockholm, Sweden
4 - Division of Neurogeriatrics, Department of Neurobiology, Care Sciences and Society, Karolinska Institute, Sweden

OBJECTIVE

Here presented are the psychometric studies so far performed on the Mindmore cognitive test battery: comparative validity against the traditional paper-based versions of the tests, test-retest reliability, usability, and a large normative study.

INTRODUCTION

Cognitive impairment is a key element in most mental disorders. Its objective assessment at initial patient contact in primary care can lead to better adjusted and timely care with personalised treatment and recovery. To enable this, we designed a self-administrative cognitive test battery intended for screening purposes in primary care (Figure 1). The battery includes 22 (sub)tests covering five cognitive domains: attention and processing speed, memory, language, visuospatial functions and executive functions.

Figure 1: Screenshot of test administration

MATERIAL & METHODS

Comparative validity was assessed in 82 participants who were administered the tests in both digital and paper-based versions (counterbalanced) with a 4-week interval. Validity was measured through Pearson's (RAVLT, PASAT and TMT A and B) and Spearman correlations (Corsi Span forward and backward, Stroop Test and BNT) and equivalence tests[1].
Test-retest reliability was assessed in another 40 participants with two test occasions and a 4-week interval. Reliability was calculated by means of the Intra-class correlation coefficient (ICC) using the two-way mixed effects model, single measurement, and absolute agreement[2].
Normative data was collected in 720 healthy adults, representing the Swedish population in terms of age span (range 17 to 93), level of education and sex (Table 1). Regression models were fitted on the data. The BIC criterium[3] was used to select the optimal predictors among the primary predictors age, years of education and sex, and secondary predictors age[2], 1/years of education, age x years of education, and age x sex. To measure usability, normative participants were asked to answer on a scale from 0 to 5 how willing they were to perform these tests again if their doctor asked them to. All tests were administered in a controlled environment with the tests presented on a 12.3" touchscreen tablet. All analyses were performed in R version 3.6.0[4]. ICCs were calculated using the package irr[5].

Table 1. *Demographics of the final normative sample*

		Age		Education		Sex	
Ages	*n*	*Mean*	*SD*	*Mean*	*SD*	Female	Male
<20	33	17.82	0.81	10.94	0.93	18	15
20 – 29	88	24.61	3.07	14.55	1.92	40	48
30 – 39	107	33.92	2.93	16.36	2.66	43	64
40 – 49	99	44.62	2.88	15.96	2.92	51	48
50 – 59	128	54.76	2.74	15.34	2.68	79	49
60 – 69	121	64.80	3.01	15.02	2.79	84	37
70 – 79	118	73.84	2.52	15.25	3.23	79	39
80+	26	84.08	4.18	13.54	3.08	20	6
Total	**720**	**50.76**	**18.78**	**15.14**	**2.93**	**414**	**306**

RESULTS

Significant correlations were observed between all digital and traditional paper-based tests (median *r*=0.53, range 0.34-0.67, Table 2). Score equivalence between test versions was observed for TMT A and B, RAVLT Learning, STR and LTR, Corsi Span forward and backward, PASAT, but not for the BNT and Stroop Test[6].
Significant ICC(A,1) were obtained for all 14 (sub)tests (median ICC=0.64, range 0.28-0.78, Table 2). A practice effect was observed for TMT parts A and B, SDPT, Corsi Span backwards, FAS and PASAT, but not for the other eight test measures. Interestingly, no practice effects were observed for RAVLT in which alternate word lists were employed.

Normative models were established for all 22 (sub)tests[7]. The test results were most affected by age and to a lesser extent by years of education and sex (Figure 2). Out of 720 normative subjects, 637 answered the usability question. Most participants indicated willingness to do the test again, with 98% providing an above average answer (≥3 on a scale from 0 to 5; Figure 3).

Table 2. Validity, reliability and normative data for 22 Mindmore cogntive (sub)tests

Cognitive domains and test	Cognitive function	normative *n*	comparative validity *r*	reliability ICC
Attention and processing speed				
Trail Making Test (TMT) part A index	mental speed	550	0.62	0.64
Symbol Digit Processing Test (SDPT)	visual scanning	199		0.75
Simple Reaction Time Test index	simple reaction time	416		0.65
Memory				
RAVLT learning	auditory learning (15 words)	372	0.53	0.78
RAVLT short-term recall	5min recall	439	0.42	0.68
RAVLT long-term recall	20min recall	437	0.59	0.68
RAVLT recognition index	20min recognition	450	0.34	0.45
CERAD learning	auditory and visual learning (10 words)	186		
CERAD recall	5min recall	186		
CERAD recognition index	5min recognition	212		
Corsi span forward	spatial short-term memory	692	0.59	0.47
Corsi span backward	spatial working memory	680	0.53	0.28
Language				
Token Test	auditory comprehension	183		
Boston Naming Test (BNT-15)	naming	536	0.67	
Letter Fluency Test FAS	verbal fluency	184		0.76
Visuospatial functions				
Cube Drawing Test	figure copying	467		
Clock Drawing Test	figure drawing	164		
Executive functions				
TMT part B index	mental flexibility	554	0.53	0.52
Complex Reaction Time Test index	complex reaction time	395		0.56
Paced Auditory Serial Addition Test (PASAT)	information processing	256	0.43	0.34
Stroop Test index	concentration effectiveness	552	0.66	0.63
Tower of Hanoi index	planning	209		

RAVLT = Rey Auditory Verbal Learning Test (15 words); CERAD = Word List Learning Test (10 words). Index measures correspond to accuracy / speed[8]. Comparative validity correlations were calculated on total completion time, resp. reaction time, for TMTA, TMTB and the Stroop Test.

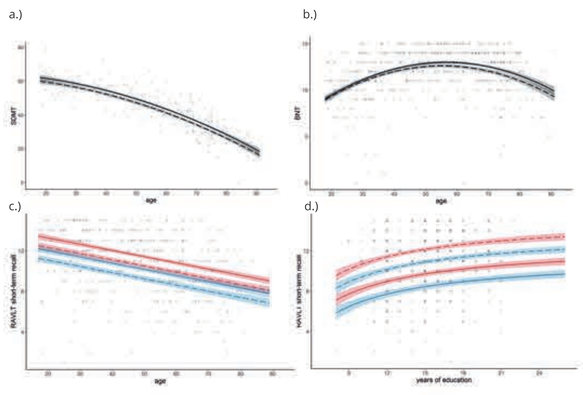

Figure 2: Examples of regression-based normative models

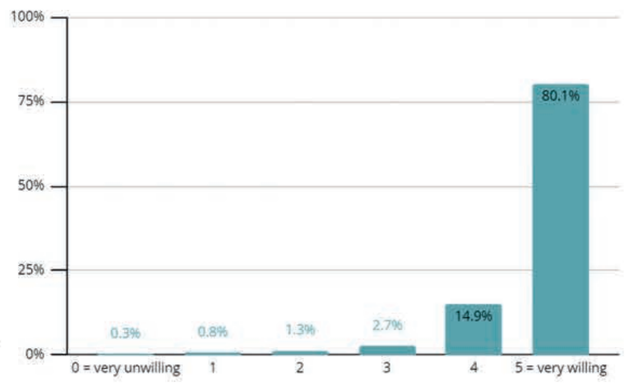

Figure 3: How willing are you to take these tests again if your doctor asked you to? (n=632)

a. SDPT showed an accelerated decline with age and a positive effect of more years of attained education; b. BNT displayed a U-shaped age effect with optimal performance around 55 years and a positive effect of more years of attained education; c. and d. RAVLT Short-term recall showed a linear decline with increasing age, a female advantage, and a non-linear positive effect of more years of attained education with the largest effect for the first 12 years of education. Line colour indicates sex: red=female; blue=male; line solidity in panels a-c indicates years of education: dashed=12 years; solid=16 years; line solidity in panel d indicates age: dashed=35 years; solid=75 years; shaded area=confidence interval.

DISCUSSION

The degree of compliance between the digital and paper-based test versions showed that they were, at least moderately, correlated and that test scores, for those tests with small modifications, were statistically equivalent. Thus, demonstrating the concordance between the digital and analogue test versions. Average to good reliability was found for ten out of the 14 tests included, showing that these tests have the characteristics to be used with multiple testing sessions. Lower reliability for RAVLT Recognition, Corsi Span and PASAT indicate that these tests are less suitable for repeated testing. No practice effect was observed for eight of the tests, whereby it was confirmed that alternate wordlists for RAVLT results in no test-retest effects. Comparison to normative data provides a measure of standardised deviation of a patient's observed result to their expected result had they been part of the healthy normative population. A large majority of participants indicated willingness to take the tests again if their doctor would ask them to.
To further determine the clinical validity of the digital test versions, we shall follow up with studies on sensitivity in different clinical populations. Currently, data collection is ongoing in the following patient populations: dementia, MCI, depression and clinical burn-out.

CONCLUSION

The psychometric properties for the Mindmore test battery appear as following:

- **Decent comparative validity to the traditional cognitive tests**
- **Good reliability for most tests, with lower reliability for some**
- **A large normative database representative of the Swedish population in terms of age, education and sex**
- **A first indication for very high usability at least among healthy subjects**

The presented psychometric properties allow clinicians to accurately interpret their patients' test results obtained through cognitive screening, hopefully leading to improved clinical decision making and better care for patients with cognitive impairment.

REFERENCES

1. Lakens, D. Equivalence Tests. *Soc. Psychol. Personal. Sci.* **8**, 355–362 (2017).
2. McGraw, K. O. & Wong, S. P. Forming Inferences About Some Intraclass Correlation Coefficients. *Psychol. Methods* **I**, 30–46 (1996).
3. Chowdhury, M. Z. I. & Turin, T. C. Variable selection strategies and its importance in clinical prediction modelling. *Fam. Med. Community Heal.* **8**, e000262 (2020).
4. R Core Team. R: A Language and Environment for Statistical Computing. (2019).
5. Gamer, M., Lemon, J., Fellows, I. & Puspendra, S. irr: Various Coefficients of Interrater Reliability and Agreement. (2019).
6. Björngrim, S., Van den Hurk, W., Betancort, M., Machado, A. & Lindau, M. Comparing Traditional and Digitized Cognitive Tests Used in Standard Clinical Evaluation – A Study of the Digital Application Minnemera. *Front. Psychol.* **10**, 1–11 (2019).
7. Van den Hurk, W., Bergman, I., Machado, A., Bjermo, J. & Gustavsson, A. Swedish Normative Data for Mindmore: A Comprehensive Cognitive Screening Battery, Both Digital and Self-Administrated. *J. Int. Neuropsychol. Soc.* 1–15 (2021) doi:10.1017/S135561772100045X.
8. Thorne, D. R. Throughput: A simple performance index with desirable characteristics. *Behav. Res. Methods* **38**, 569–573 (2006).

More about our research: https://www.mindmore.com/evidens