

PSYCHOMETRIC PROPERTIES OF THE TOKEN TEST

Yulia Akinina^{1,2}, Olga Buivolova¹, Olga Soloukhina¹, Anastasia Shlyakhova¹,
Roelien Bastiaanse^{1,2}

¹National Research University Higher School of Economics, Moscow, Russia

²University of Groningen, Groningen, the Netherlands

Rome, September 24, 2019

OUTLINE

- A brief overview of the Token Test
- The Token Test App
- Psychometric properties of the Token Test App (so far)

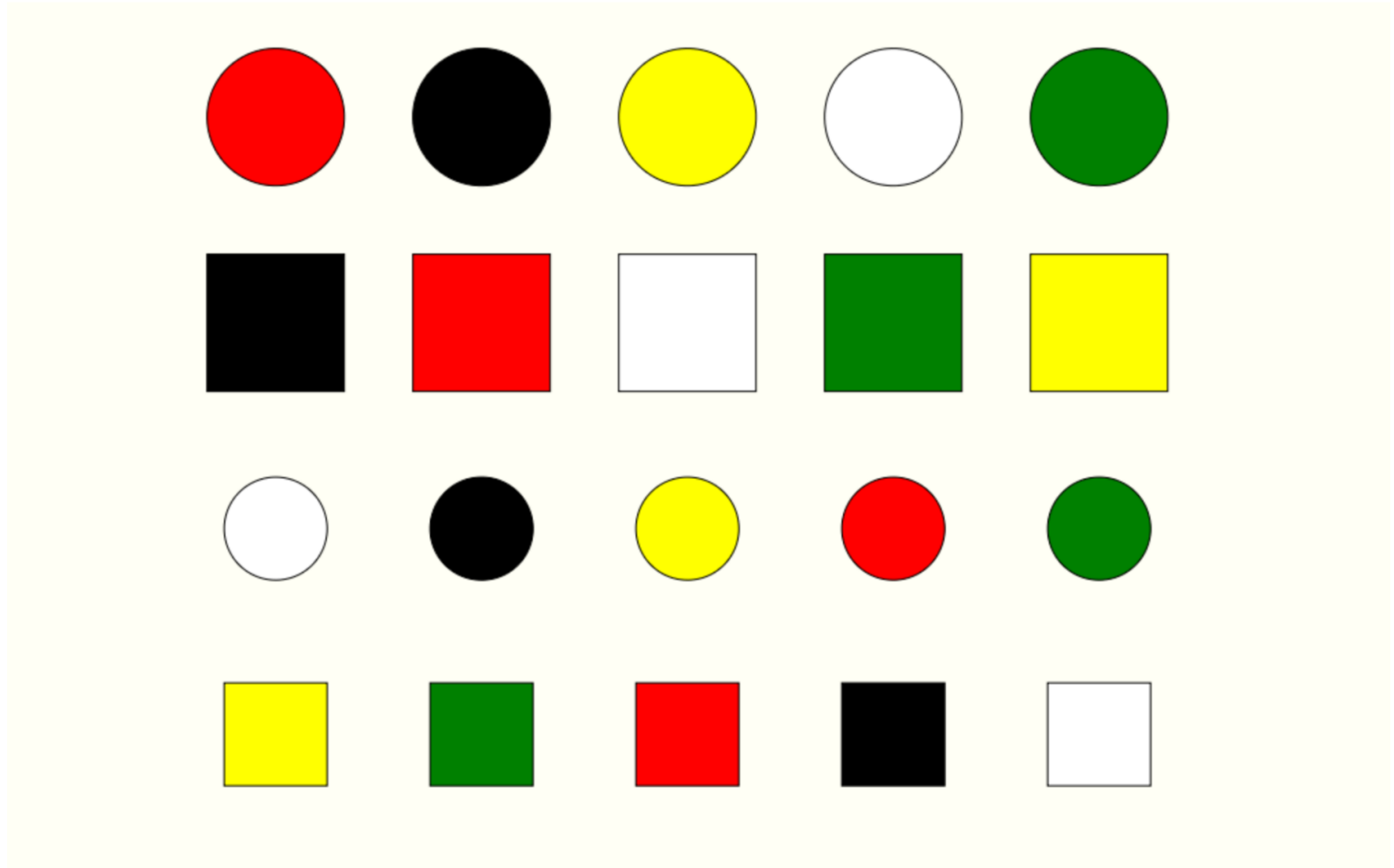
THE TOKEN TEST: A BRIEF OVERVIEW

The Token Test

- De Renzi, E. and Vignolo, L.A. (1962) The Token Test: A sensitive test to detect receptive disturbances in aphasics. *Brain*, 85, 665-678.
- Tokens of different shape, color and size are presented to the participant, and they follow the instructions of varying complexity
- Idea: you have to understand every single word
- Measures auditory comprehension, independent of intelligence
- Presence and severity of aphasia in general
- Many variants and applications

The Token Test: shortened version

- De Renzi E., Faglioni P. Normative data and screening power of a shortened version of the Token Test // Cortex. 1978. Vol. 14. P. 41-49.
- 36 instructions in six blocks
- Increasing complexity
 - Block 1: “Touch a red token”
 - Block 5: “Touch the large white circle and the little green square”
 - Block 6: ”Touch all the circles except for the green one”
- Correct response = 1, correct response after repetition = 0.5, incorrect response = 0 (max total score = 36)



”As you can see, there are twenty tokens here ...“ (de Renzi & Faglioni, 1978)

THE TOKEN TEST APP

Advantages of computerized testing

- Reduction of human error
- Standardization of the procedure
- Automatic presentation and scoring
- Time and financial efficiency

(Newton et al., 2013)

The first edition

- Aim: build an electronic version of the Token Test for the tablet (eTT)
 - Multiple languages
 - Available to everyone
- Bastiaanse R. Raaijmakers S., Satoer D. The e-Token Test. 2015. Groningen (NL): Groningen Expert Center for Language and Communication Disorders
- Promising psychometric properties (Akinina et al., 2017), but **unusable** app

The second edition

- Presentation and scoring bugs fixed
- Timing corrected
- Visual feedback
- Discontinue conditions according to (De Renzi & Faglioni, 1978) original paper
- New languages
- Each language checked by back-translation
- iOS and Android operating systems

Languages available

- Afrikaans
- Akan
- Albanian
- Armenian
- Berber
- Bosnian
- Catalan
- Catalan from Valencia
- Chinese Mandarin (Mainland)
- Chinese Mandarin (Taiwan)
- Croatian
- Czech
- Danish
- Dutch
- English (American)
- English (Australian)
- English (British)
- English (Canadian)
- English (South African)
- Finnish
- Flemish
- French
- Frisian
- Galician
- German
- Greek
- Hebrew
- Hungarian
- Maltese
- Norwegian
- Persian
- Portuguese
- Portuguese (Brazilian)
- Russian
- Spanish
- Swiss German
- Tagalog
- Tatar
- Turkish

PSYCHOMETRIC PROPERTIES OF THE TOKEN TEST APP

MAIN COHORT

100 NHI

Various age and level of education

Data on tablet experience (1-6 scale)

100 PWA

Various aphasia types and severity

Data on tablet experience (1-6 scale)

Effects of demographic variables and tablet experience

Differences between NHI and PWA

Differences within the PWA group

Diagnostic properties: cut-off scores, sensitivity & specificity



RELIABILITY

TEST-RETEST RELIABILITY

NHI group:

- 20 participants
- Two weeks between two sessions

PWA group:

- 20 participants
- Two weeks between two sessions
- No SLP in between

VALIDITY

CONCURRENT VALIDITY

NHI and PWA groups

20 participants

eTT vs paper-and-pencil TT version

CONCURRENT VALIDITY-II

PWA group

eTT vs standard clinical severity assessment

MAIN COHORT

109 NHI

Various age and level of education
Data on tablet experience (1-6 scale)

100 PWA – *in progress*

Various aphasia types and severity
Data on tablet experience (1-6 scale)

Effects of demographic variables and tablet experience

Differences between NHI and PWA

Differences within the PWA group

Diagnostic properties: cut-off scores, sensitivity & specificity



RELIABILITY

TEST-RETEST RELIABILITY

NHI group:

- 20 participants
- Two weeks between two sessions

PWA group: - *in progress*

- 20 participants
- Two weeks between two sessions
- No SLP in between

VALIDITY

CONCURRENT VALIDITY

NHI and PWA groups – *in progress*

20 participants each
eTT vs paper-and-pencil TT version

CONCURRENT VALIDITY-II

PWA group - *in progress*

eTT vs standard clinical severity assessment

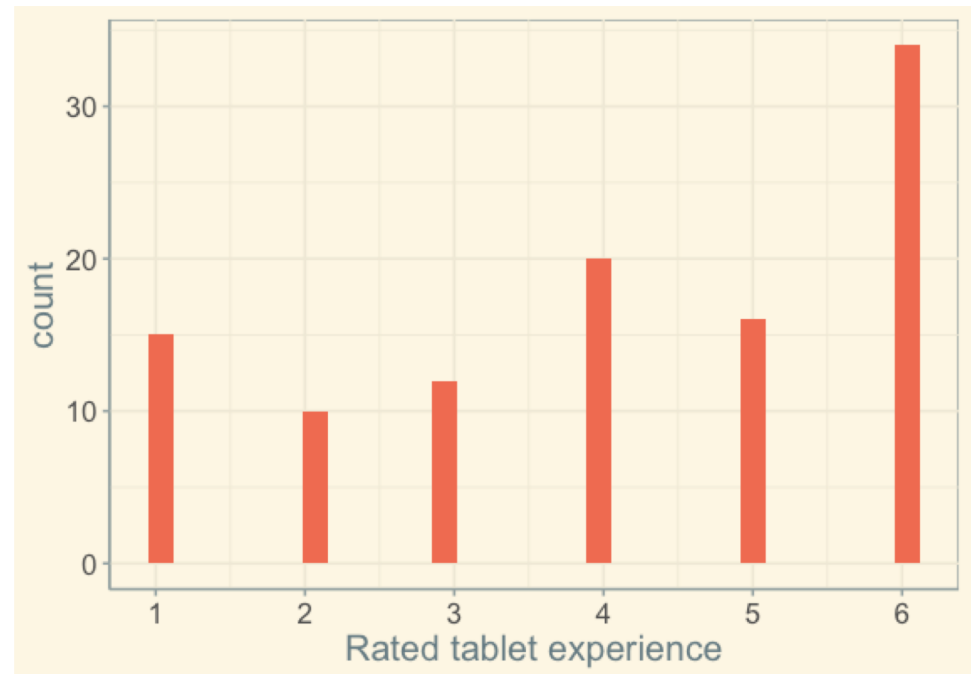
Main cohort: NHI

- 109 neurologically healthy Russian speakers tested with eTT version 2.6.0
- 64 female
- Mean age 41 years old (18-75, $SD = 18.39$)
- Mean 15 years of education (10-25, $SD = 2.55$)

		Age				
Education (years)		18-30	31-50	51-70	70+	(sum)
10-12	female	6	1	1	1	9
	male	5	1	1	0	7
13-14	female	6	3	2	3	14
	male	9	0	3	0	12
15-16	female	6	11	8	0	25
	male	5	5	3	1	14
17+	female	4	6	3	3	16
	male	2	3	6	1	12
	(sum)	43	30	27	9	109

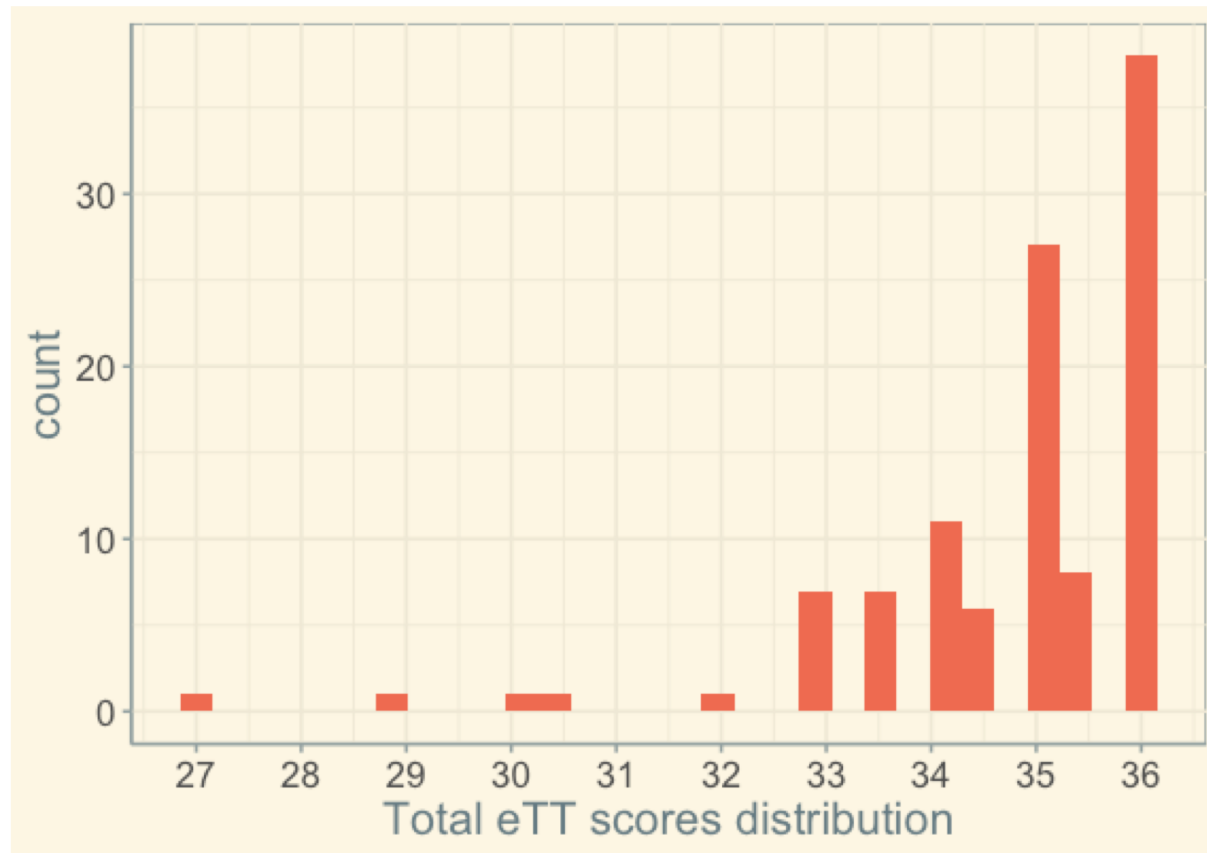
Main cohort: NHI

- Tablet experience (data missing for two participants)
- Self-rated degree of experience with tablets on a scale from 1 (never used a tablet) to 6 (use it every day)
- Mean tablet experience of 4.07 (1-6, $SD = 1.79$)



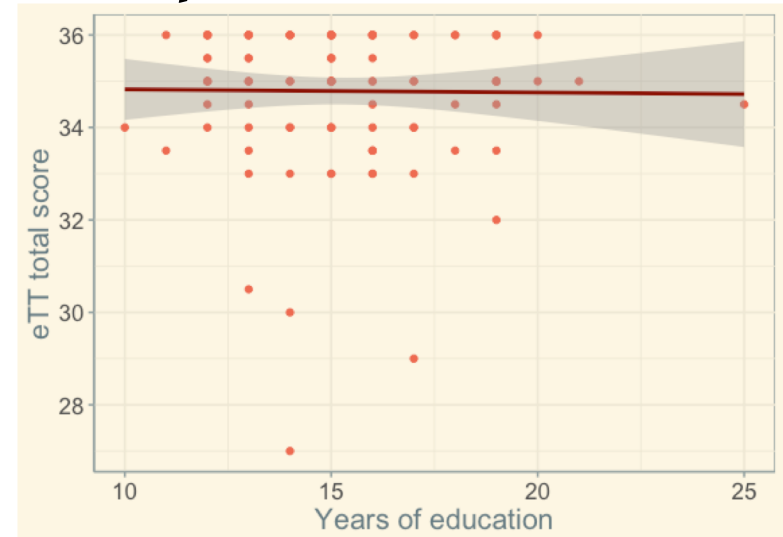
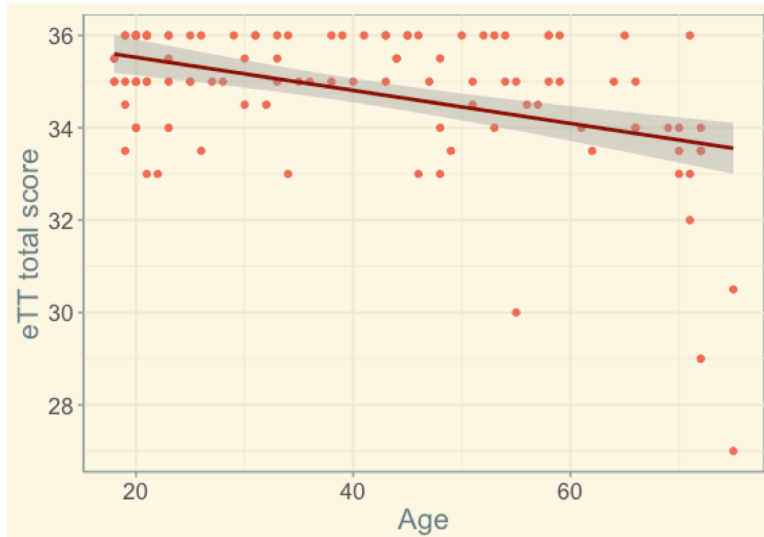
Main cohort: NHI - eTT results

- Mean total score = 34.79, Median = 35 (27 – 36, $SD = 1.51$)



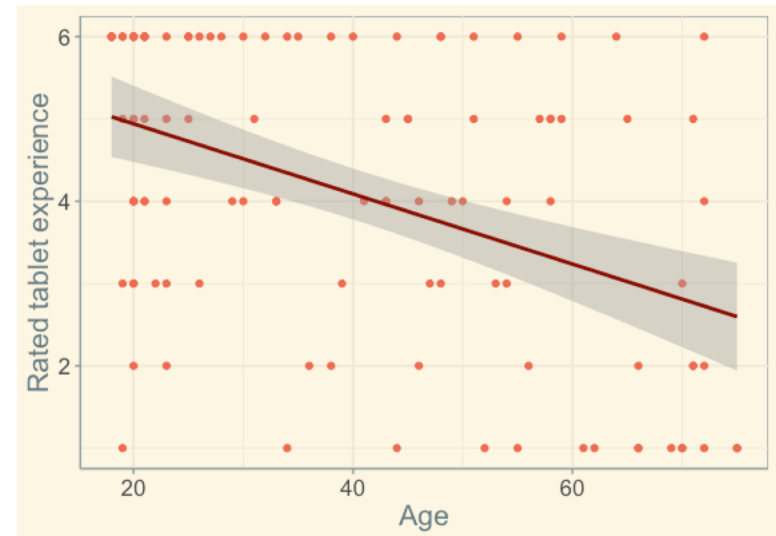
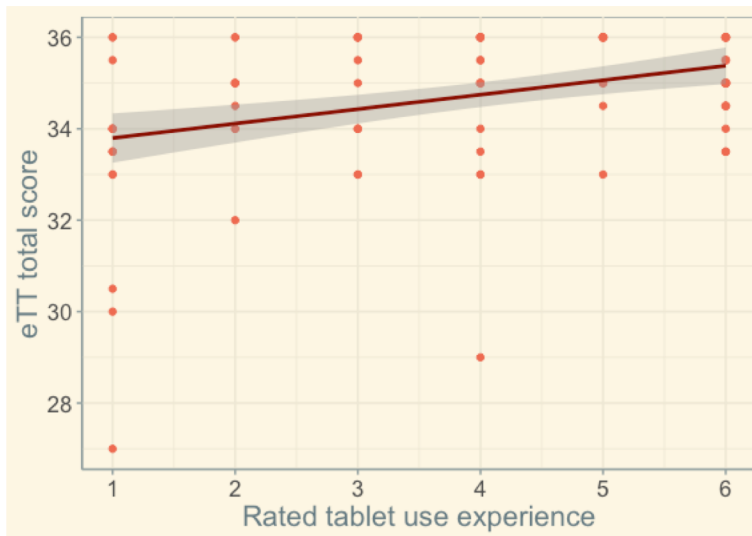
Effects of demographic variables

- Males ($M = 34.82$) VS Females ($M = 34.77$): no significant effects of **sex**
- Correlation analysis:
 - Significant negative moderate correlation with **age** ($r = -.44$, $p < .0125$)
 - No correlation with **education** level ($p = .9$)



Effects of tablet experience

- Correlation analysis:
 - Significant positive moderate correlation with tablet experience ($r = .37, p < .0125$)
- However:
 - Significant correlation between age and tablet experience ($r = -.44, p < .0125$)



Interim summary

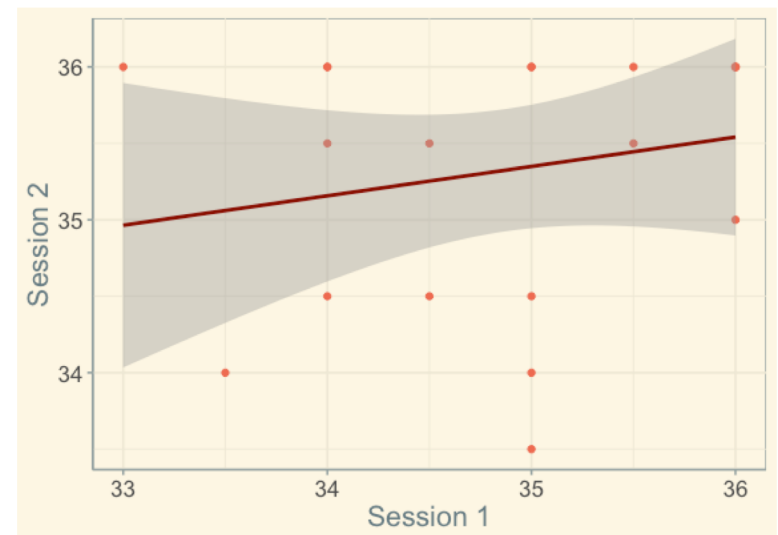
- Sex and level of education do not correlate with the eTT total score
- Age and tablet experience correlate with the eTT total score
- Regression analysis is needed to derive the adjusted TT score formula that takes into account relevant variables
 - The best way to take into account the tablet experience?

Test-retest reliability: NHI

- 20 Russian-speaking NHI, 13 females, mean age 42 years old (20-72, $SD = 17.16$), mean 16 years of education (12-25, $SD = 3.04$), mean tablet experience 3.8 (1-6, $SD = 1.32$)
- Tested twice (average time between sessions ~ 14 days)
- Test-retest reliability:

correlation **non-significant**

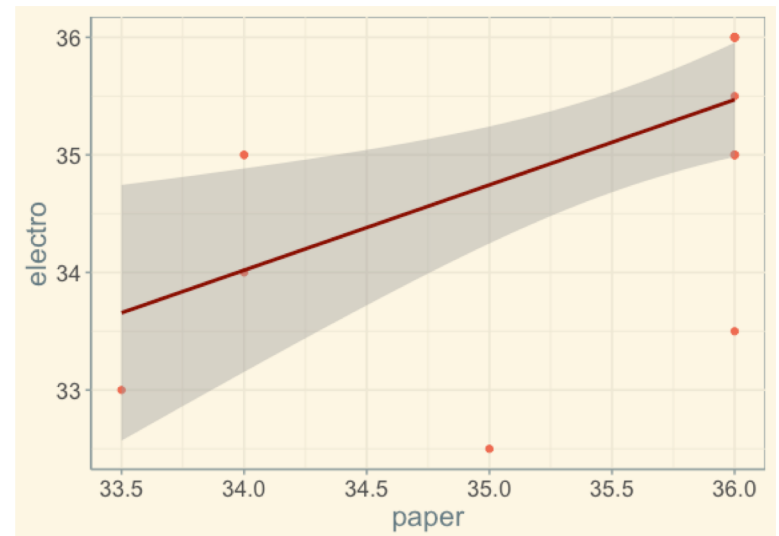
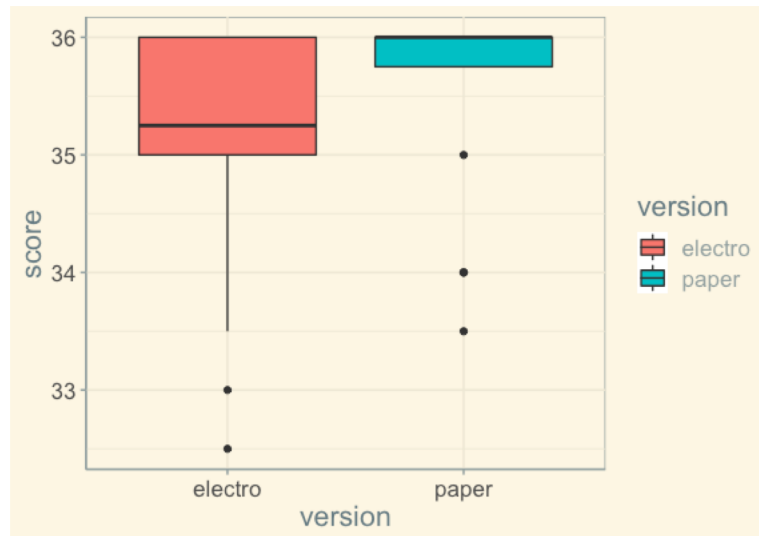
- Practice effects:
 - Average gain 0.45 (-1.5 - 3)
 - Paired-sample one-sided permutation test **non-significant**



(Akinina et al., 2019)

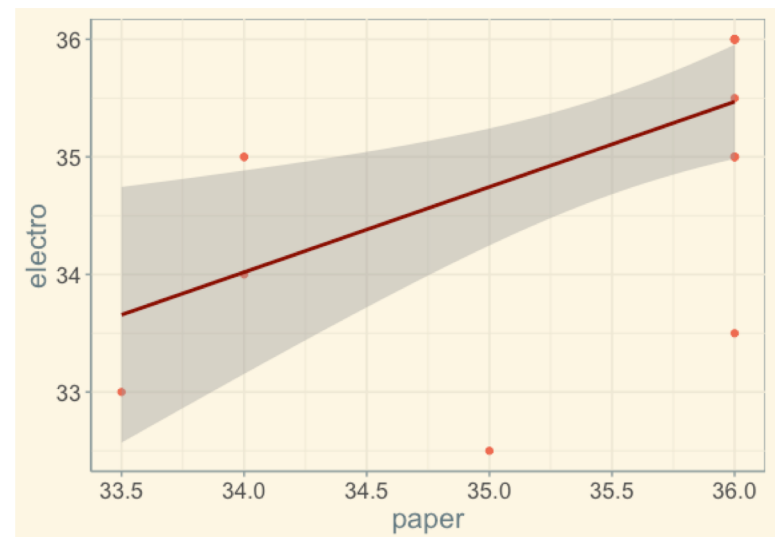
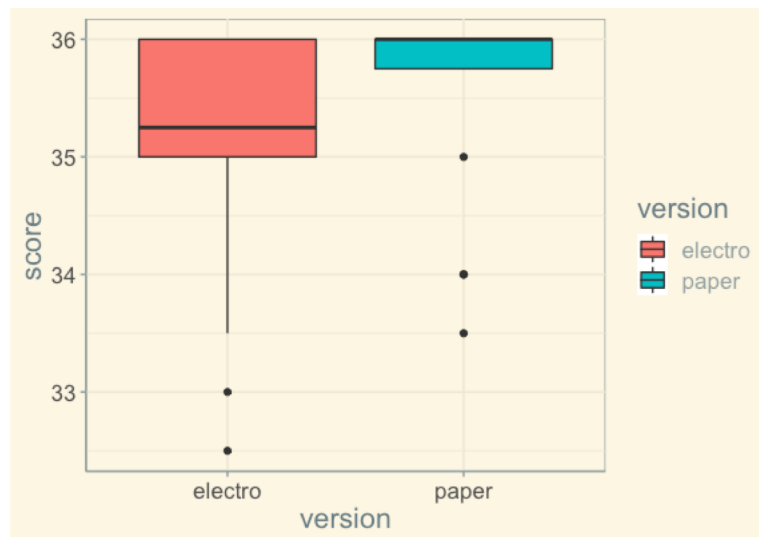
Concurrent validity-I: NHI

- 20 Russian-speaking NHI, 13 female, mean age 46 years old (21 – 70, $SD = 16.24$), mean 15 years of education (10-22, $SD = 2.33$), mean tablet experience of 3.8 (1-6, $SD = 1.51$)
- Paper-and-pencil and eTT versions, order counterbalanced between participants



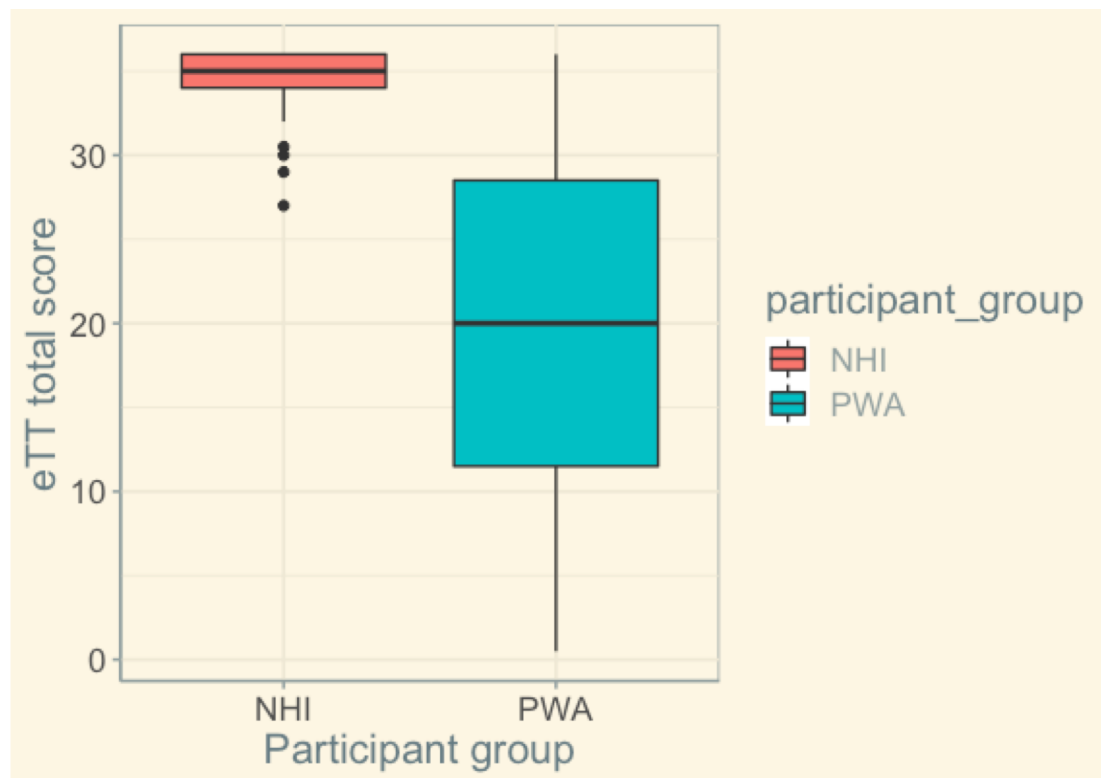
Concurrent validity-I: NHI

- Mean difference (paper – eTT total score) = 0.4 (-1 – 2.5, $SD = 0.91$)
- The difference between versions is **non-significant** (a paired permutation test, $p = .09$)
- Correlation analysis: **significant** positive correlation (spearman ranked test w/permutations, $p = .0011$)



PWA data: a sneak peek

- 67 PWA with subacute and chronic aphasia of various etiology (mostly stroke) and aphasia type
- Mean score of 19.32 (0.5 – 36, $SD = 10.21$)



SUMMARY AND FUTURE WORK

Summary

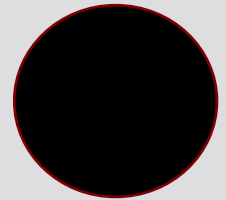
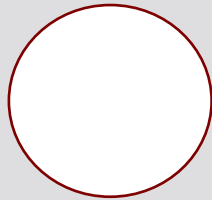
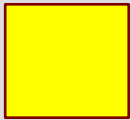
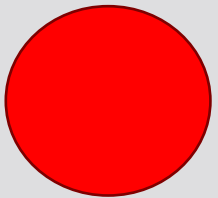
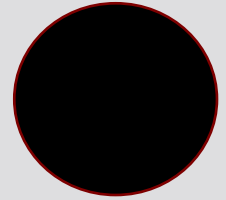
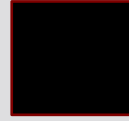
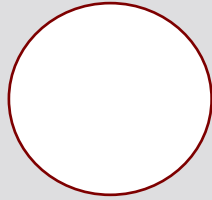
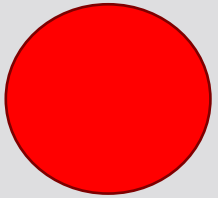
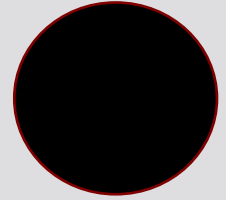
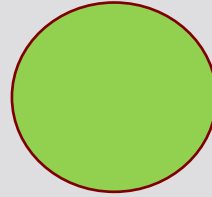
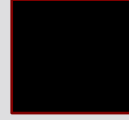
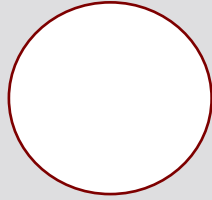
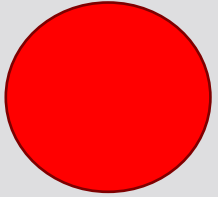
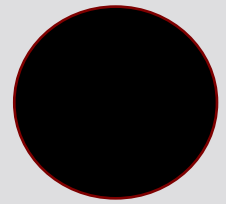
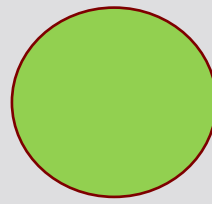
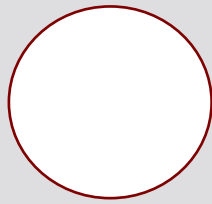
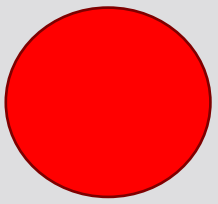
- Age and tablet experience affect the eTT score and need to be adjusted for
- The test-retest reliability could not be demonstrated in the NHI group, probably due to the ceiling effect
 - The practice effect is negligible
 - Needs to be verified in the PWA group
- The eTT version demonstrates good concurrent validity in the NHI group
 - Needs to be verified in the PWA group

Future work: the next steps

- Derive the adjusted TT score formula based on the NHI data
- Conclude the collection of the PWA data
- Analyze test-retest reliability and concurrent validity in PWA
- Evaluate the diagnostic properties of the eTT: cut-off scores, sensitivity & specificity

Future work: multilingual transfer?

- Many languages available
 - Are norms and cutoffs calculated for one language (Russian) applicable to other languages?
- General rule: normative data have to be collected for each one! (Ivanova & Hallowell, 2013)
- A very resource-intensive task
- Is there another way?



Thank you!

Questions?