



Automated Language Deficit Diagnostics with the Token Test App

Test-Retest Reliability and Practice Effects

INTRODUCTION

- Aphasia is an acquired language disorder due to brain damage in adults. It is one of the most common and devastating consequences of stroke. Aphasia manifests in difficulties speaking and comprehending speech, reading and writing.
- According to some estimations, 30% of first-ever ischemic stroke survivors will have aphasia (Engelter et al., 2006), and reliable diagnosis of aphasia is crucial for an adequate rehabilitation plan. Token Test (de Renzi & Vignolo, 1962) is the most widely used screening diagnostic test in clinical practice.
- Digital versions of language diagnostic tools have many advantages compared to their paper-and-pencil counterparts, such as reduction of human error, standardization of the procedure, automatic scoring and time and financial efficiency (Newton et al., 2013). This is why a multilingual electronic version of the Token Test (eTT) was developed.

THIS STUDY aims to investigate *test-retest reliability* and *practice effects* of the eTT in a group of neurologically healthy individuals. Test-retest reliability indicates measurement stability across time and is calculated by correlating the scores from the same participants tested on several occasions. Practice effects are calculated as significant gains in scores upon test repetition. Additionally, we examine the *influence of tablet experience* on the test performance.

THE TOKEN TEST

- Many versions exist
- 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.
- 20 tokens of different shape, color and size are presented to the participant, and they follow instructions of varying complexity
- 36 instructions organized in six blocks
- Either touching or manipulating the tokens
- Correct response = 1, correct response after a single repetition = 0.5, incorrect response = 0 (total score range = 0 - 36)
- Measures auditory comprehension
- Indicates presence and severity of aphasia in general

LANGUAGES CURRENTLY AVAILABLE

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

PARTICIPANTS AND PROCEDURE

- 20 neurologically healthy Russian-speaking adult individuals
- Tested with the eTT Version 2.6 in Russian on two occasions
- Average time between the two sessions = 14.05 days (range 12-16, SD = 1)
- Average tablet experience = 3.8 (range 1-6, SD = 1.32) on a scale from 1 ("I never used one") to 6 ("I use it every day")

TABLE I.
TOTAL SCORES DESCRIPTIVE STATISTICS

	mean	median	SD	min	max	IQR
session 1	34.88	35.0	0.92	33.0	36.0	1.63
session 2	35.33	35.75	0.85	33.5	36.0	1.5
session 2 - session 1	0.45	0.25	1.11	-1.5	3.0	1.0

INSTRUCTION EXAMPLES:

- Block 1 (all tokens): Touch a circle
Touch a yellow token...
- Block 2 (large tokens): Touch the yellow square
Touch the black circle...
- Block 3 (all tokens): Touch the small white circle...
- Block 4 (large tokens): Touch the red circle and the green square...
- Block 5 (all tokens): Touch the large white circle and the small green square...
- Block 6 (large tokens): Put the red circle on the green square
Put the green square next to the red circle
Touch all the circles, except the green one...

RESULTS AND DISCUSSION

Test-Retest Reliability. The correlation between the total scores in the two sessions was not significant. This indicates low test-retest reliability for eTT. However, note that in both sessions the participants performed almost at ceiling, and the score range was very narrow (see Table 1). Additionally, six out of twenty participants had zero score difference between the sessions. **Practice Effects.** The practice effects were relatively low (lower than the lowest test point gain; mean = 0.45, median = 0.25) and non-significant in neurologically healthy individuals. If eTT scores of a person with aphasia improve on consecutive testing sessions, this would be likely due to improvement in language ability rather than practice effects. **Tablet Experience and eTT Performance.** The correlations between the rated tablet experience and the total scores in the first session, as well as between the across-sessions total score differences and the rated tablet experience, were not significant.

CONCLUSIONS AND FUTURE WORK

eTT is suitable for tracking language improvement over time. No significant practice effects were observed, and test performance, as well as its change over time, was not related to the rated experience with tablets. The test-retest reliability obtained in our study was low: a measurement error exists, but its clinical relevance is unclear. It might be inappropriate to measure test-retest reliability in participant groups where variation in the level of ability is not expected, any variation in the test score and covariance in the scores between sessions are probably random. An additional test-retest reliability study in people with aphasia is necessary. This study is a part of a project on standardization of eTT and establishing its' psychometric properties.

ACKNOWLEDGEMENT We would like to thank Anastasia Shlyakhova for her help with data collection.

Yulia Akinina^{1,2} jakinina@hse.ru, Roelien Bastiaanse^{1,2}, Olga Soloukhina¹

Center for Language and Brain¹/Neurolinguistics dep., Faculty of Arts²

National Research University Higher School of Economics¹ (Moscow, Russia)/University of Groningen² (Groningen, the Netherlands)