Estonian vocabulary test LexEst correlates with self-assessed proficiency

Kaidi Lõo (kaidi.loo@ut.ee)[1], Katrin Leppik[1], Anton Malmi[1], Agu Bleive[1], Raymond Bertram[2]

[1] University of Tartu, [2] University of Turku

Vocabulary knowledge is a good predictor of language proficiency among L2 speakers across CEFR levels (A1 to C2) and is a prime indicator of reading comprehension (Milton, 2010; Laufer & Ravenhorst-Kalovski, 2010). Essentially, measuring vocabulary size offers profound insights into an individual's language capabilities in understanding and communication (Staehr, 2008). Consequently, an accurate vocabulary assessment tool for L2 speakers is valuable both for scientific research and practical applications.

In recent years, various versions of the *Lexical Test for Advanced Learners of English* (*LexTALE*, Lemhöfer and Broersma, 2012) have been developed to measure vocabulary in various languages. While the original English test was intended for advanced language learners, subsequent variants of this test (e.g., *LexTale-Esp* for Spanish, *LexIta* for Italian, *Lexize* for Finnish) have shown that successful vocabulary tests can be developed for L2 learners at different proficiency levels (Izura et al. 2014, Amenta et al. 2021, Salmela et al. 2021). The current study introduces LexEst, marking the first instance for a similar quick vocabulary assessment tool in Estonian.

As other vocabulary tests, LexEst makes use of a visual lexical decision task (cf. Lõo et al. 2018) in which participants decide whether a letter string is an existing Estonian word or a nonword. The current study was run online with 135 items (90 words & 45 nonwords) and 500 participants from 18-77 years old, of which 274 were native speakers and 226 non-native speakers). Most non-native speakers were Russians (n=116), followed by Germans (n=25) and Finns (n=16).

We assessed the quality of items using point-biserial correlation and item response theory analysis considering difficulty and discrimination rates (Rizopoulus, 2006). By excluding the easiest items and items of low discriminability, we decreased the original item list from 135 to 90 items (60 words & 30 nonwords). For Estonian L2 learners, the vocabulary score on the final set of items correlated well with self-assessed proficiency (r=0.73). There was also a medium correlation between vocabulary score and duration of stay in Estonia (r=0.51) and daily language use of Estonian (r=0.37). In conclusion, the LexEst vocabulary test enables a quick and objective assessment of L2 learners' vocabulary in Estonian, that resonates well with more subjective measures of language proficiency such as self-assessed language skills.

References

Amenta, S., Badan, L., & Brysbaert, M. (2021). LexITA: A quick and reliable assessment tool for Italian L2 receptive vocabulary size. *Applied Linguistics*, *42*(2), 292-314.

Izura, C., Cuetos, F., & Brysbaert, M. (2014). Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicológica*, *35*(1), 49-66.

Laufer, B. & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited. *Reading in a foreign language, 22*, 15-30.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*, 325-343.

Lõo, K., Järvikivi, J., & Baayen, R. H. (2018). Whole-word frequency and inflectional paradigm size facilitate Estonian case-inflected noun processing. *Cognition*, *175*, 20-25.

Milton, J. (2010). Development of vocabulary breadth across CEFR levels. In Bartning et al.(Eds). Eurosla monograph (pp. 211-232).

Rizopoulos, D. (2006). ltm: An R Package for Latent Variable Modeling and Item Response Analysis. *Journal of Statistical Software*, *17*(5), 1–25.

Salmela, R., Lehtonen, M., Garusi, S., & Bertram, R. (2021). Lexize: A test to quickly assess vocabulary knowledge in Finnish. *Scandinavian Journal of Psychology*, *62*(6), 806-819.

Staehr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal, 36,* 139-152.