Applied Psycholinguistics (2025), **46**, e39, 1–26
doi:10.1017/S0142716425100301

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL ARTICLE

Introducing LexEst: a quick and efficient vocabulary test for assessing vocabulary knowledge in L2 Estonian

Kaidi Lõo[1], Katrin Leppik[1], Anton Malmi[1], Agu Bleive[1] and Raymond Bertram[2]

[1]University of Tartu, Tartu, Estonia and [2]University of Turku, Turku, Finland
Corresponding author: Kaidi Lõo; Email: kaidi.loo@ut.ee

(Received 20 January 2025; revised 9 July 2025; accepted 23 September 2025)

Abstract

This study presents the creation and validation of LexEst, a short 5-minute test to assess vocabulary knowledge in Estonian. Our freely accessible test consists of 90 items and is designed for L2 speakers of Estonian. LexEst is modeled after the original Lexical Test for Advanced Learners of English. Similarly to other test variants, our test has been adapted to assess vocabulary knowledge at varying proficiency levels. Our findings demonstrate that LexEst provides an objective measure of the Estonian vocabulary of L2 learners, aligning well with subjective language proficiency indicators, such as self-assessed skills. In addition, higher LexEst scores and shorter response times are associated with higher CEFR-level language courses and a greater daily use of Estonian. Higher LexEst scores are also associated with an earlier age of acquisition in Estonian and a higher perceived importance of learning Estonian.

Keywords: Estonian; language proficiency; lexical decision; response times; vocabulary

Introduction

Estonian is a Finno-Ugric language of the Uralic language family, with approximately 1.1 million native speakers (L1). Furthermore, according to the 2021 census,[1] a considerable part of the population in Estonia (about 17%) speaks Estonian as a second language (L2). This number is expected to grow in the coming years as Estonia transitions from a country with a negative net migration rate to one with a steady inflow of migrants who need to learn Estonian as an L2. Additionally, Estonian schools are in transition from Russian and Estonian medium education to only Estonian medium education. With the increasing number of Estonian L2 speakers comes the need for researchers and educators to find means to measure Estonian language proficiency reliably. However, at the moment, there is no quick and easy way to achieve this. The current study introduces the Estonian version of

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0142716425100301 Published online by Cambridge University Press

the LexTALE vocabulary knowledge test (Lemhöfer & Broersma, 2012). LexEst (https://lexest.ut.ee/) is a freely available vocabulary test for speakers of Estonian as a second language. It offers a quick and easy way to measure learners' proficiency through vocabulary assessment.

So far, Estonian language skills, including vocabulary, have been measured via CEFR-level testing[2] in schools and by the Education and Youth Board. Regarding freely available assessment options, the Estonian Integration Foundation[3] offers online screening and diagnostic tests, allowing language learners to assess their Estonian knowledge by themselves. They include, among other skills tests, a vocabulary assessment test. The Institute of the Estonian Language has also developed several tools[4] to aid teachers with Estonian as a second language (Kallas et al., 2021; Üksik et al., 2021), including modules on vocabulary, grammar, communicative language activities, and text evaluation. Vaks et al. (2025) are currently developing a test battery including sentence repetition, non-word repetition, and a vocabulary test that enables an objective assessment of language development in bilingual children.

It is important to recognize, however, that most of the resources for the general audience as the ones discussed above are quite comprehensive and time-consuming to conduct and evaluate, which is typically not well suited for use in a research setting where the time available for testing a participant is mostly limited and the language level assessment is not the focus of the study. Such tests are generally not specifically designed for scientific purposes and are generally not validated or tested for reliability. In conclusion, there is no objective, efficient, and reliable test to be used in a research setting that assesses the language proficiency of adult Estonian L2 speakers, neither in general nor specifically regarding their vocabulary knowledge.

To address the restrictive-time issue, it has been common practice in scientific studies to use self-ratings and language background questionnaires (e.g., LEAP-Q) to measure L2 proficiency. There is no doubt that they provide useful information and are generally considered quite reliable (cf. Kaushanskaya et al., 2019; Marian et al., 2007), yet may not fully align with objective performance measures, such as standardized test scores and grades (Zell & Krizan, 2014), or are at least less consistent than more objective measures such as vocabulary tests. One reason may be the variation in questionnaires and inconsistencies in the way self-ratings are collected (see Grosjean, 1998), with respect to the scale being used. Another reason could be related to differences in how individuals or groups of individuals rate themselves when using self-rating scales (Grosjean, 1998). For instance, younger individuals may be more likely to overrate themselves by using extreme scale points, whereas older individuals often provide more conservative evaluations. However, it is not so that self-ratings come from nowhere, that is, they typically do capture language proficiency skills as well. To illustrate this, Lemhöfer & Broersma (2012) found that error rates and response times of their English vocabulary test LexTALE correlated well with scores from other tasks such as a translation test and another lexical decision test. However, self-ratings also showed correlations with these tasks, though to a slightly lesser extent. Interestingly, in the same study, self-ratings correlated better with the Quick Placement Test results than the vocabulary test scores did. This suggests that self-ratings and vocabulary scores largely tap into overlapping constructs, but also capture some distinct aspects, making them both

https://doi.org/10.1017/S0142716425100301 Published online by Cambridge University Press

valuable. Therefore, the use of objective measures, such as a vocabulary test, would at least provide a valuable supplement to subjective measures and improve the accuracy and reliability of language proficiency assessments in research settings.

Although proficiency in a second language involves more than simply knowing many words (see Read, 2000), a growing body of research indicates that evaluating lexical competence in L2 provides a strong indication of overall proficiency in L2. Zareva et al. (2005) showed that vocabulary size is more strongly related to L2 learners' proficiency than other linguistic skills. Milton (2010) reviewed the findings of a series of common vocabulary tests. He found that vocabulary size predicts language proficiency among L2 speakers across the six proficiency levels of the Common European Framework of Reference for Languages (CEFR levels A1 to C2). Participants with higher proficiency had higher scores on vocabulary tests. It is also worth noting that vocabulary size relates to proficiency across the four language modalities (writing, reading, speaking, listening; see, e.g., Laufer & Ravenhorst-Kalovski 2010; Milton 2010; Stæhr 2008; Zimmerman 2004). From these four modalities, Stæhr (2008) and Milton (2010) found that writing and reading are the abilities that are most closely related to the size of the vocabulary. Miralpeix & Muñoz (2018) analyzed the relationship between vocabulary size and all four skills and reached a similar conclusion: vocabulary size is more closely related to writing and reading than to speaking and listening. At the same time, a systematic meta-analysis of 100 individual vocabulary studies (Zhang & Zhang, 2022) confirmed a strong relationship between knowledge of the L2 vocabulary and L2 reading and listening skills. One challenge is that, according to the overview, more than 80% of vocabulary studies focus on English as L2. To our knowledge, the relationship between vocabulary knowledge and proficiency in Estonian as an L2 has not been investigated.

The findings on the relationship between vocabulary, L2 proficiency, and response times have been more inconclusive. Laufer & Nation (2001) and Harrington (2006) found that L2 speakers with a higher level of proficiency (intermediate vs. advanced) not only scored higher but were also faster on a vocabulary test. In contrast, Miralpeix & Meara (2014) reported no differences in response times in vocabulary tests between intermediate and advanced second language speakers. In the current study, we not only assess the accuracy rate to evaluate the extent of L2 speakers' vocabulary but also use response time as an indicator of the time it takes to activate words in the mental lexicon. A proficient L2 speaker should not only know a lot of words but also be able to access them fast enough to communicate efficiently.

Overview of existing vocabulary tests in other languages

Given the dominance of English in second language research, most vocabulary tests have been developed for English as a second language. A commonly used test is the Vocabulary Levels Test (VLT, Nation 1983). The Vocabulary Levels Test is a multiple-choice test with four difficulty levels, each level including words from a different frequency band. At each level, tasks are presented in groups of six words, and the learner must match three of them with the three provided definitions (e.g., *pencil* => something used for writing). The test assesses participants'

https://doi.org/10.1017/S0142716425100301 Published online by Cambridge University Press

language proficiency based on the number of correctly identified items at various difficulty levels. It has proven to be a valuable tool for assigning students to appropriate ability groups in language classes.

Another vocabulary test, the Eurocentres Vocabulary Size Test (EVST, Meara & Jones 1988), is divided into 10 blocks, each including words from a different frequency band to measure vocabulary knowledge. Instead of matching the meaning to form, the test uses a visual lexical decision task in which L2 learners must decide whether a string of letters represents a real English word (e.g. *weekday*) or a nonword (e.g. *klimp*), in other words, a letter string that resembles an English word but does not actually exist in English. Unlike the Vocabulary Levels Test, the Eurocentres Vocabulary Size test attempts to measure the absolute size of a learner's vocabulary.

A lexical decision task was also utilized in the Lexical Test for Advanced Learners of English (LexTALE), developed by Lemhöfer & Broersma (2012). Compared to VLT and EVST, LexTALE is considerably shorter. This 60-item test includes a carefully selected mix of 40 low to high frequency words as well as 20 pseudowords. Thus, also here, the task is to choose between a yes and no answer to indicate whether the letter string on the screen is an English word or not. Lemhöfer & Broersma (2012) validated their test with Korean and Dutch advanced learners of English. They found that LexTALE test scores are better predictors of other proficiency tasks, including L1-L2 translation tasks, than self-rating proficiency. Thus, their results indicated that this kind of test can be used with learners of different L1 backgrounds. Lemhöfer & Broersma (2012) also developed vocabulary tests for L2 Dutch and German, but these tests have not been validated.

As the English test was shown to be easy to use and a good indicator of English L2 proficiency, in the following years, similar tests have been created for other languages. For example, tests were created for French (Lextale FR, Brysbaert 2013), Spanish (LextaleESP, Izura et al. 2014), Chinese (LEXTALE CH, Chan & Chang 2018 and LexCHI, Wen et al. 2024), Italian (LexITA, Amenta et al. 2021), Portuguese (LextPT, Zhou & Li 2022), Finnish (Lexize, Salmela et al. 2021), Malay (LexMAL, Lee et al. 2024), and Arabic (LexArabic, Alzahrani 2024). Subsequent variants of this test have shown that a reliable vocabulary test can be developed for L2 learners of varying proficiency levels, not just advanced learners. In general, test scores differentiate well between L1 and L2 speakers, as well as between L2 speakers' different CEFR levels (e.g., Amenta et al. 2021). This, in turn, indicates that the test gives a good approximation of the general proficiency level of a learner.

All subsequent studies in languages other than English also showed that there are significant correlations between test scores and scores on other assessments, for instance, proficiency self-ratings of L2 learners for reading, listening, writing, and speaking. The test scores of LexCHI, LexMal, and LexTALE also correlate well with a widely used cloze test (see e.g., Wilson, 1953), where participants have to fill out text passages with gaps, e.g., "Today, I went to the __ and bought some milk and eggs." The cloze test measures participants' ability to understand vocabulary in context. Finally, Lextale FR found a correlation between L2 participants' test scores and the number of years they have learned French, and the Finnish Lexize found a correlation between participants' test scores and the time they had been exposed to Finnish.

https://doi.org/10.1017/S0142716425100301 Published online by Cambridge University Press

For L1 learners, the test has typically been too easy, and the scores do not reflect individual differences well enough (Amenta et al., 2021; Brysbaert, 2013; Dijkgraaf et al., 2017; Izura et al., 2014). In other words, with many of the vocabulary tests, a ceiling effect is observed as L1 learners get very high scores (Amenta et al., 2021; Izura et al., 2014; Lemhöfer & Broersma, 2012). An exception is the Finnish Lexize test (Salmela et al., 2021), which showed that older adult L1 speakers scored higher than younger adult L1 speakers. With Lexize, it was also found that for L1 speakers, higher secondary school grades were related to higher vocabulary scores and that the level of tertiary education was related to vocabulary knowledge. However, typically L2 vocabulary tests are designed to span the full range of L2 vocabulary knowledge, which means that the test is generally not suitable for detecting differences at the higher end of L1 proficiency, for instance, among students with higher tertiary education. It is reasonable to expect that many of these tests can capture differences in L1 vocabulary at lower proficiency levels or as a function of educational background (cf. Salmela et al. 2021). Moreover, they have proven suitable for use with L1 school-aged children whose vocabulary is still developing, as demonstrated in a recent large-scale Finnish study involving children aged 9 to 15 years (Bertram et al., 2025).

The current study

The goal of the current study is to create an objective and validated vocabulary test for Estonian L2 speakers that can easily be used in research settings. We consider vocabulary knowledge to be a proxy for language proficiency, and to this date, there is no quick, reliable, and easy way to assess language proficiency of adult L2 Estonian speakers for research purposes. The aim is to design a test that can be used for all proficiency levels of Estonian L2 speakers, not only for advanced Estonian L2 speakers (cf. the original English LexTALE). Our study also uses a demographic and language background questionnaire, allowing us to examine the relationship between LexEst scores and self-assessed Estonian proficiency, age of acquisition, language use and exposure, as well as CEFR level of the L2 Estonian courses participants had participated in. The use of subjective proficiency ratings to validate the vocabulary test does not imply that these ratings represent a definitive measure of proficiency. However, a positive correlation with subjective ratings suggests that the test reflects certain aspects of what learners perceive as proficiency, even if these perceptions are not entirely accurate. Subjective ratings are just one of several criteria used in the validation process, alongside factors such as nativeness (L1 vs. L2), L2 CEFR level, time spent in Estonia, age of acquisition, and frequency of Estonian use. By combining multiple indicators, we can reduce the impact of the limitations of any single measure, including self-assessments. For the vocabulary test to demonstrate validity, its scores should return a difference between L1 and L2 speakers, exhibit a positive correlation with CEFR levels and subjective proficiency ratings. Additionally, they should correlate positively with the amount of time spent in Estonia and the frequency of use of Estonian, while showing a negative correlation with the age of acquisition.

Previous Lextale-type studies usually did not collect response times, as the test was not timed. The few studies in which response times were considered showed

https://doi.org/10.1017/S0142716425100301 Published online by Cambridge University Press

that L1 speakers have shorter response times than L2 speakers and that more proficient L2 speakers are faster than less proficient L2 speakers (see Harrington 2006; Laufer & Nation 2001). In the current study, we expect to observe a similar relationship between the response times of the participants and their Estonian proficiency. However, it should be noted that because of the online nature of the current test, the results with response times should be taken with caution. Variability in internet connection speed, device responsiveness, background sounds, or other aspects of participants' testing environments may have introduced additional noise into the response time data, potentially obscuring more subtle relationships with individual language history variables.

Method

The original LexEst stimuli selection

The stimuli for the LexEst Estonian vocabulary test were selected from the Balanced Corpus of Estonian, which comprises 15 million tokens: 5 million tokens from newspaper texts, 5 million tokens from fiction, and 5 million tokens from scientific texts (Kaalep & Muischnek, 2005). The words in the corpus were divided into six frequency bands. Subsequently, 30 words were randomly selected from each of the six frequency bands.

This initial list of 180 words was further reduced to 90 words. We only included nouns, adjectives, and verbs and excluded compounds, words with obvious foreign origin,[5] possibly offensive words, and words with lexical ambiguity (e.g., *luud* means "bones" and "broom") from this initial list. We also excluded derived words that were transparent, such as *janu-ne* ("thirst-y"). Since transparently derived words usually contain more frequent morphemic constituents (e.g., *joy +ful => joyful*), test takers can guess meanings based on familiar stems or affixes rather than truly knowing the words. This can give a misleading impression of how well and how many (especially lower frequency) words are known. In other words, including transparent derived words (as well as compounds) could overestimate vocabulary knowledge, which is why we have avoided them. This approach aligns with the rationale used in the vocabulary tests developed by Brysbaert and colleagues, where transparent morphologically complex words were also avoided.

The final selection included 10 words with a frequency of less than 1 per million (pm); 24 words with a frequency of 1 to 5 pm; 22 words with a frequency of 5 to 10 pm; 18 words with a frequency of 10 to 20 pm; 10 words with 20 to 100 pm; and 6 words with a frequency greater than 100 pm. This division in frequency bands and the stimulus selection procedure followed the procedure used in other similar studies, e.g., Finnish (Salmela et al., 2021). The length of the stimuli varied between 3 and 8 characters. We included 50 nouns, 24 adjectives, and 16 verbs, reflecting the proportions of these word classes as they roughly occur in the Estonian Balanced Corpus. Nouns make up about 60%, verbs about 10%, and adjectives around 17% of all word classes in the Estonian Balanced Corpus.

The nouns and adjectives were all in nominative singular form, and the verbs were in *ma*-infinitive form. The selection of 45 pseudowords was based on a 45 word shadow list from the same corpus. The shadow list was a list of Estonian words

https://doi.org/10.1017/S0142716425100301 Published online by Cambridge University Press

that mimicked the length and frequency distribution of the word list. Pseudowords were created by changing 1–3 letters in existing Estonian words from this shadow list. The length of the pseudowords varied between 3 and 7 characters. We excluded pseudowords that were real words in other major foreign languages (e.g., *rana* "frog in Spanish"). Also, Estonian pseudowords that can possibly be words in slang (e.g., täägima "to tag") or in Estonian dialects (e.g., *aader* "arter") were not included.

Participants

L1 participants were recruited through social networks and university email lists. L2 participants were recruited inside and outside Estonia via social media and targeted emails to language schools and teachers. L1 speakers in our study are individuals who acquired Estonian as their first language and use it as their primary language of communication. L2 speakers are individuals who began acquiring Estonian at age two or later and do not use Estonian as their main language of communication in daily life.

Altogether, 533 participants took part in the study, but data from 48 participants were removed after initial screening. More specifically, data for nine participants were removed as they did not complete the whole study, and data for three participants were removed as they were under 18 years old. Data from one participant were removed because the age was not indicated, and data from a second one because the indicated age was in contradiction with other questions in the questionnaire. Six participants were removed because they stated they were dyslexic. Finally, 28 speakers who reported acquiring Estonian from birth, i.e., they were raised in environments where at least one caretaker (parent) primarily used Estonian, but another language was also spoken from birth, were excluded from the analysis, as they did not clearly fit into either the L1 or L2 group. The data from the remaining 485 participants (18–77 years, mean 39.6 years, 348 females and 129 males, 8 non-binary) were analyzed. The data included 274 L1 speakers and 211 L2 speakers of Estonian. The majority of L2 speakers were L1 Russian ($n = 110$), followed by German ($n = 26$), Finnish ($n = 17$), and English ($n = 11$). Table 3 in Appendix B shows the whole diverse distribution of native languages across our L2 speakers. In our study, the age of acquisition (AoA) for Estonian as an L2 ranged from 2 to 70 years (mean 19.9 years). Notably, unlike most Lextale-type studies, which typically set a later minimum AoA, we intentionally included an earlier AoA to gain a more comprehensive understanding of its impact on vocabulary acquisition and to ensure that the test would be suitable across a wide range of L2 proficiency levels.

Language background questionnaire

In addition to the LexEst test, we asked participants to complete a language background questionnaire. We asked participants about their age, gender, education, daily Estonian use, the age they started learning Estonian (age of acquisition), how important learning Estonian is for them (1—not important, 10—very important), and to rate their own language proficiency in listening, speaking, writing, and reading (1—very weak, 10—outstanding). In addition to their self-

https://doi.org/10.1017/S0142716425100301 Published online by Cambridge University Press

Table 1. Distribution of L2 speakers across proficiency levels

L2 proficiency level	Number of L2 speakers
A1	26
A2	22
B1	31
B2	38
C1	40
Total	157

assessed proficiency, we also asked the participants to indicate the highest CEFR level of the Estonian language course they had participated in (A1—beginner to C2—advanced). One hundred fifty-seven L2 speakers reported their Estonian language class level, 54 left the question unanswered. Table 1 shows that the distribution of participants across course levels was relatively equal, with most participant from the B level. While a higher CEFR level typically reflects greater language proficiency, it does not necessarily indicate stronger motivation to learn Estonian. For instance, learners enrolled in C1-level courses are likely motivated to achieve advanced proficiency in Estonian because of personal goals or work-related demands. However, learners at the B-level may also be highly motivated, for example, in pursuit of the B1-level diploma required for Estonian citizenship.

Finally, participants were asked to indicate the device they used to complete the task and the location where they completed it (at home/at work/in a classroom in Estonia/abroad). In other words, the experiment was done outside the lab without the presence of an experimenter, similar to the procedure in other Lextale-type vocabulary tests. The statistical significance of these factors was evaluated using Kruskal–Wallis tests. Neither the location of the participants (home, work or classroom, or whether they were in Estonia or abroad, nor the device used (computer, tablet, or mobile phone) had a significant effect on the results.

Procedure

The study was carried online using the learning platform ViLLE,[6] a platform developed at the University of Turku, Finland.

Participants first completed a background questionnaire, followed by the LexEst vocabulary test. The vocabulary test makes use of a visual lexical decision task in which participants decide whether a letter string is an existing Estonian word or a pseudoword (cf. Lõo et al. 2018). There were six practice trials and 135 experimental trials. Each trial started with a fixation point for 500 ms, after which a letter string appeared on the screen for six seconds. The participants had to decide whether the letter string was a word or not by pressing either the *yes*- or the *no*-key. If they did not make a decision within six seconds, the experiment automatically moved on to the next trial. Participation was voluntary, and the entire procedure took about 10 minutes.

https://doi.org/10.1017/S0142716425100301 Published online by Cambridge University Press

Item assessment

After collecting the data, we limited our item set to select the best items for the final item set, following the same procedure as was used in other Lextale-like studies (e.g., Lextale FR, Brysbaert 2013). Item assessment was done using item response theory in the R-package *ltm* (Rizopoulos, 2007).

We assessed our items (90 words and 45 pseudowords) for accuracy, difficulty, and discriminability as well as goodness-of-fit. We first assessed the quality of items using point-biserial correlation and item response theory analysis. L1 and L2 participant groups were evaluated together. Words and pseudowords were assessed separately. The point-biserial correlation was calculated between each subject's mean accuracy for items and accuracy for a given item. A positive point-biserial correlation coefficient indicates that participants who performed well on the overall test tended to answer a particular item correctly, while those who performed poorly tended to answer it incorrectly.

The point-biserial correlation coefficient ranges between +1 and −1. The correlation for all items was positive, ranging from 0.07 to 0.78 for words and from 0.33 to 0.71 for pseudowords. Following Crocker & Algina (1986), we set the correlation threshold to 0.2 for removing items that were inaccurate. This resulted in the removal of seven words from the item list but no pseudowords.

Next, we ran two-parameter logistic models on word and pseudoword items. Whereas point-biserial correlation gives us an indication of the difficulty of an item, the IRT model provides information on both difficulty and discrimination of items. Figure 1 illustrates the difficulty and discrimination of three LexEst items. The difficulty of the items is presented on the x-axis, and the steepness of the curve represents the discriminability. In Figure 1, *kõrvits* "pumpkin" is an example of a very easy item, *väi* "son-in-law" is an example of a low discriminability item, and *niru* "lousy" is an example of an item that is both discriminative and relatively difficult.

The difficulty index shows how difficult it is to achieve a 0.5 probability of a correct response for a specific item, given the overall performance in the test. The discrimination index indicates the ability of an item to differentiate between participants who did the test with a high score from participants who did not. A good test should contain items that cover the full range of difficulty and have items with high discriminative power.

We followed Izura et al. (2014) and ordered the items based on difficulty and discrimination. We set the difficulty threshold to −1.5 and the discrimination threshold to 1. This resulted in the removal of eight words with low difficulty and one word with a low discrimination value. Two pseudowords were removed due to low difficulty, and no pseudowords were removed due to low discrimination.

Finally, following vocabulary tests in other languages and to make the test quick and easy to administer, we further limited our final test item set to 60 words and 30 pseudowords. Thus, we also removed words and nonwords based on the goodness-of-fit test in the R-package *ltm*. For 13 words, the test was significant, indicating that these were not good test items. This left us with 62 word test items. For six nonwords, the goodness-of-fit test was significant, indicating these were not good test items. This procedure left us with 37 nonword test items.

https://doi.org/10.1017/S0142716425100301 Published online by Cambridge University Press

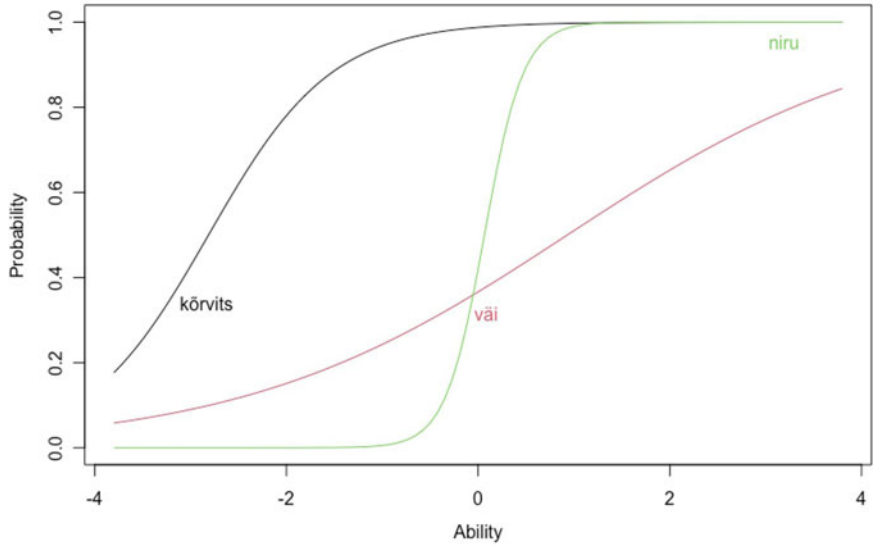


Figure 1. Item characteristic curves for *kõrvits* "pumpkin," *väi* "son-in-law," and niru "lousy".

We further removed one word that was on the lower side of the discriminability (*kärama* "to make noise" 1.4) and a word (*risu* "trash") for which there was a word with practically the same stem (*riisuma* "to rake"). Next, five relatively low discriminability pseudowords from different difficulty ranges and two pseudowords were removed to equal the percentage of verbs in the pseudoword shadow list (the list of words from which pseudowords were derived) compared to the word list.

The final set of items in the LexEst test had a high internal consistency reliability, returning a Cronbach's alpha of .96 when using the *cronbach.alpha* function in the *ltm*-package. That indicates that the final set of items in the vocabulary test consistently assess Estonian vocabulary knowledge.

The mean difficulty of the final set of items was –0.5 for words and –1.27 for pseudowords, which indicates that items, in general, when considering both L1 and L2 speakers, were not extremely difficult and that pseudowords were less difficult than words. The mean discrimination rate of 2.60 for words and 2.04 for pseudowords indicates good discriminability for the test items. The final set of words and pseudowords is listed in Appendix A. The data and the code for the item assessment are available in: https://osf.io/y42xv/.

Table 2 lists the numeric lexical characteristics of the final version of the LexEst test. Lemma frequency was calculated based on the 15 million token Balanced Corpus of Estonian (Kaalep & Muischnek, 2005) and scaled to 1 million. Bigram frequency was calculated based on the 0.5 million token Phonetic Corpus of Estonian Spontaneous Speech (Lippus et al., 2021). Bigram frequency refers to the frequency with which adjacent pairs of letters (bigrams) occur together in the corpus. We normalized each word's bigram frequency by dividing it first by its length and then by 1000.

Mean bigram frequency and length were not significantly different ($p > .05$) between words and pseu- dowords were not significantly different.

https://doi.org/10.1017/S0142716425100301 Published online by Cambridge University Press

Table 2. Lexical characteristics of the final LexEst item set

	Words (n = 60) M	SD	Range	Pseudowords (n = 30) M	SD	Range
Length in characters	4.78	1.15	3–8	4.97	1.07	3–7
Bigram frequency	22.36	9.36	4.30–40.01	22.89	8.92	2.06–39.72
Lemma frequency	15.13	35.06	0.20–265.45	–	–	–

Words from all six frequency bands were on the final list: 1 word from the first frequency band, 4 words from the second band, 18 words from the third band, 16 words from the fourth band, 17 words from the fifth band, and 4 words from the sixth band. More items from the middle frequency range and fewer from the high frequency range (1st and 2nd bands) and low frequency range (6th band) is similar to the distribution in other Lextale-like tests. The final set of items included 19 adjectives, 30 nouns, and 11 verbs.

The proficiency level of our test word items based on the teacher's tools of the Institute of Estonian Language[7] was in accordance with the frequency band distribution. Three words belonged to the A1 category, 2 words to the A2 category, most of the test items belonged to the B1 or B2 (18 words in each) proficiency category, and 6 words to the C1 category. Thirteen words were not listed by the teacher's tool.

Scoring the test

The LexEst test was scored by the following formula derived from the formula suggested by Lemhöfer & Broersma (2012) and Brysbaert (2013):

$$\text{LexEst Score} = \text{N Yes to words} - 2 * \text{N Yes to pseudowords}$$

This scoring method ensures that the guessing behavior in pseudowords is penalized. Since there are twice as many pseudowords as real words, each mistake on a pseudoword results in a 2-point deduction. A maximum score of 60 can only be obtained by saying *yes* to all words and *no* to all pseudowords. In any other case, the score will be below 60. For example, if a participant answered 40 words correctly and 10 pseudowords incorrectly (responding to them with yes), the score would be calculated as follows: 40 - 2*10 = 20. For the original LexTALE, Lemhöfer & Broersma (2012) tried out three measures, of which this one was the one that performed best (see also Brysbaert, 2013 for the argumentation). A similar formula was also used in the Italian, Chinese, Spanish, Malay, Portuguese, and Finnish tests.

Analysis

For the final set of items, we performed a series of statistical analyses using the R-package *stats* (R Development Core Team, 2024). We first compared the LexEst scores and response times of L1 and L2 speakers. We then further zoomed in on the L2 speakers and assessed how their LexEst score and response time correlated with

https://doi.org/10.1017/S0142716425100301 Published online by Cambridge University Press

their self-ratings of proficiency, the level of the Estonian class they participated in, as well as with other variables extracted from the language background questionnaire. We used the Welch two-sample t-test to test the significant differences between categorical variables with two levels (e.g., L1 vs. L2) and the Kruskal–Wallis test to test the significance between categorical variables with more than two levels (e.g., language course level). Post hoc tests for the nonparametric Kruskal–Wallis test were made with Dunn's multiple comparisons test using the R-package *dunn.test* (Dinno, 2024).

Pearson's and Spearman's correlation coefficients were used to test the significance of numeric variables (e.g., self-assessed proficiency, daily Estonian use, and age of acquisition) and word frequency. The significance of each variable was tested individually, as they were highly correlated with each other. We refrained from using more sophisticated statistical analyses such as mixed-effects regression models, as our main aim was to assess how well the vocabulary test functions and how reliable it is. As Shmueli (2010) point out, explanatory analyses are better suited for this kind of purpose than more complex predictive models, which are typically used when the goal is to predict outcomes rather than understand the underlying structure. The data and the code for the analysis are available in: https://osf.io/y42xv/.

Results

LexEst scores of L1 and L2 speakers

As illustrated in Figure 2, L1 speakers ($M = 56.58$, SD $= 5.69$, range 21–60) had a significantly higher average LexEst score than L2 speakers ($M = 23.26$, SD $= 15.65$, range -7–60). This difference was statistically significant ($M_{L1} = 56.58$, $M_{L2} = 23.26$; Welch-corrected two-sample t (251.44) $= 29.40$, $p < .0001$) with a large effect size (Cohen's $d = 2.95$). The L1 speakers performed mainly at the ceiling level. In fact, we did not find any significant relationships between the L1 score and their language background or other variables; therefore, the following analysis will focus on the L2 group.

Correlations between LexEst scores and self-assessed proficiency of the L2 speakers

Participants were asked to assess their reading, writing, listening, and speaking proficiency on a scale from 1 (low) to 10 (high). In the analysis, there was a similar and high correlation between all components of proficiency. The LexEst score correlated positively with self-assessed ability to speak (Pearson's $r = .74$, $p < .0001$), listen ($r = 0.72$, $p < .0001$), read ($r = .73$, $p < .0001$), and write ($r = .74$, $p < .0001$) in Estonian. Participants who rated their proficiency in these four modalities higher also achieved higher scores on the LexEst test. This is illustrated in Figure 3.

Correlations between LexEst scores and Estonian language course level of L2 speakers

Next, the Kruskal–Wallis test showed a significant effect of the proficiency level of Estonian courses ($H = 78.79$, $df = 4$, $p < .001$). Dunn's multiple comparison test

https://doi.org/10.1017/S0142716425100301 Published online by Cambridge University Press

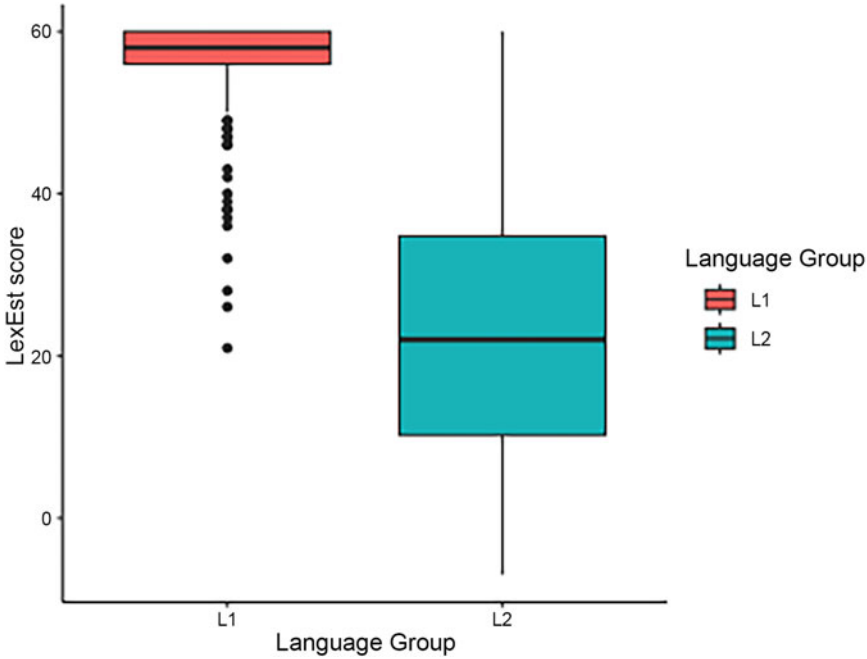


Figure 2. Boxplot illustrating the distribution of LexEst scores in the L1 and L2 groups.

revealed no significant difference between participants the A1 and A2 course levels ($Z = -0.65$, $p = .26$) and between participants in the B1 and B2 course ($Z = -0.44$, $p = .33$). Therefore, we merged together levels A1 and A2 as a single course level A and levels B1 and B2 together as a single course level B. The multiple comparison test with new levels showed that participants in the course level A scored significantly lower than participants from both course level B ($Z = -3.45$, $p = .0003$) and course level C ($Z = -7.61$, $p < .0001$), and that participants in course level B in turn scored significantly lower than level C ($Z = -3.89$, $p = .0001$).

Figure 4 shows that there was a significant gradual relationship between the course level and the score.

A1-level participants scored the lowest, and C1-level participants the highest on the test.

Correlations between LexEst scores and other participant variables

There was a significant correlation between the LexEst score and the participant's daily language use of Estonian ($r = .52$, $p < .0001$). The more participants spoke Estonian daily (on the scale from 0–100 percent), the higher their LexEst score. There was also a correlation between the LexEst score and participants' duration of stay (from 0 to 61 years) in Estonia ($r = .37$, $p = .0001$), the longer participants had lived in Estonia, the higher their score. Furthermore, there was a correlation between the LexEst score and the participant's age of acquisition ($r = -.42$, $p = .0001$), i.e., the age when they started learning Estonian (from 2 to 67 years).

https://doi.org/10.1017/S0142716425100301 Published online by Cambridge University Press

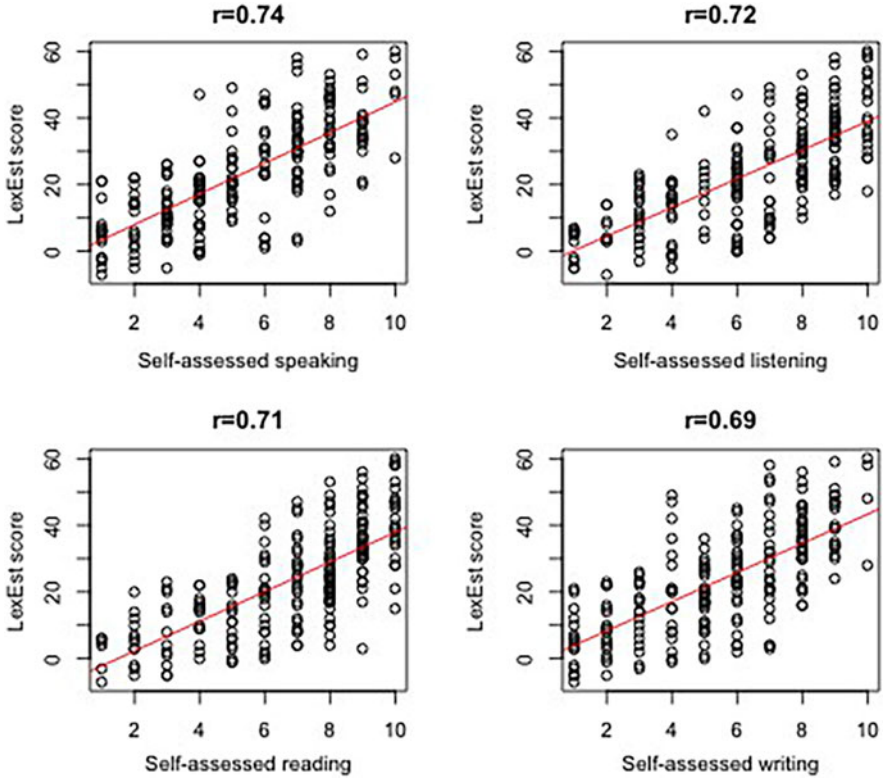


Figure 3. Correlations between LexEst scores and self-accessed proficiency of L2 speakers across four modalities.

Finally, there was a correlation between the LexEst score and the importance of speaking Estonian ($r = .30$, $p = .0001$). That means participants who valued learning Estonian more (on a scale of 0—not important at all to 10—very important) also scored higher on the test. All these correlations are illustrated in Figure 5.

LexEst response times

Finally, we analyzed the response times of the final set of LexEst word items. First, incorrect responses were removed (2082 data points, 11% of the data). In addition, responses longer than four seconds were identified as outliers. We decided to keep the response time cutoff relatively long to accommodate potential variability introduced by the online testing environment. We removed them from the data set (832 data points, 3% of the data).

As illustrated in Figure 6, L2 speakers (M = 1.55 s, SD = 0.71 s, range 0.42-3.99s) were slower than L1 speakers (M = 1.07 s, SD = 0.43 s, range 0.35–3.96 s). This difference was statistically significant ($M_{L1} = 1.07$, $M_{L2} = 1.55$; Welch-corrected two-sample $t\,(10213) = 54.53$, $p < .0001$) with a large effect size (Cohen's

https://doi.org/10.1017/S0142716425100301 Published online by Cambridge University Press

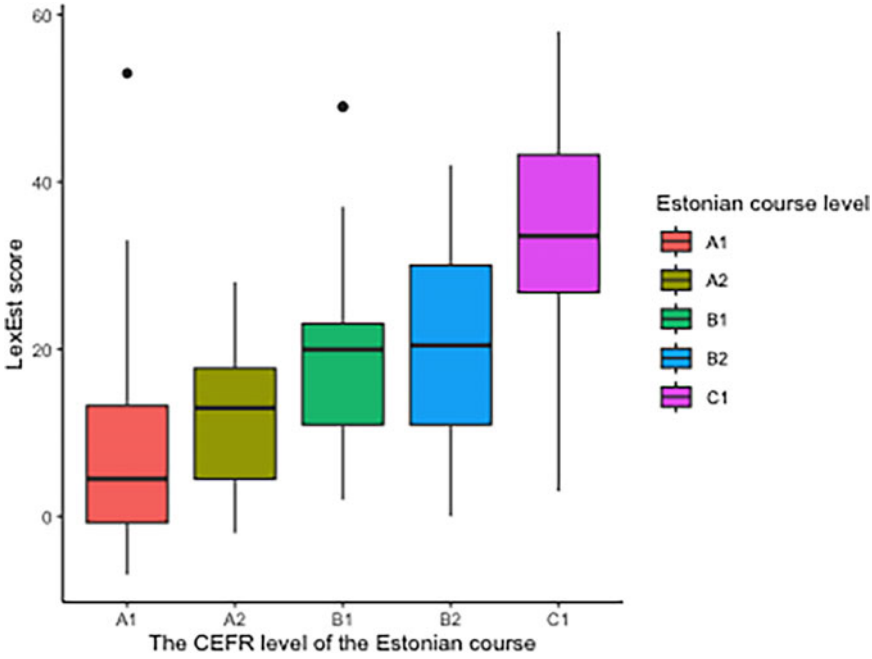


Figure 4. Boxplot showing the distribution of LexEst scores among CEFR levels of L2 speakers.

$d = 0.90$). Similarly to LexEst scores, there were almost no significant correlations between language background and response times for native speakers. The only significant correlation was between age and response time. We found that older participants were slower than younger participants ($r = .17$, $p < .0001$, cf. Allen et al. 1991).

Similarly to LexEst scores, L2 response times correlated with self-assessed language proficiency scales. However, the correlations for response times were smaller than those for LexEst scores. L2 speakers tended to respond faster when they reported higher self-assessments of their reading (Pearson's $r = -.18$, $p < .0001$), listening (Pearson's $r = -.21$, $p < .0001$), writing (Pearson's $r = -.14$, $p < .0001$), and speaking abilities (Pearson's $r = -.21$, $p < .0001$). The analysis also showed that L2 speakers who used Estonian more daily were faster than L2 speakers who used Estonian less ($r = -.19$, $p = .001$).

Furthermore, the participant's language course level also correlated with response times (see Figure 7). The Kruskal–Wallis test showed a significant effect of Estonian course level ($H = 169.5$, $df = 4$, $p < .0001$). The Dunn's multiple comparison test showed that there were no significant differences between language course levels B1 and B2, so we decided to merge them together as one level. Participants who had participated in the A1 level course were significantly slower compared to participants who had participated in the A2 level language course ($Z = 4.15$, $p < .0001$), in B level ($Z = 6.22$, $p < .0001$) and in C1 level language course ($Z = 7.33$, $p < 0.0001$). Participants with A2 level course were also significantly slower than participants with B level course ($Z = 5.33$, $p < = .0001$)

https://doi.org/10.1017/S0142716425100301 Published online by Cambridge University Press

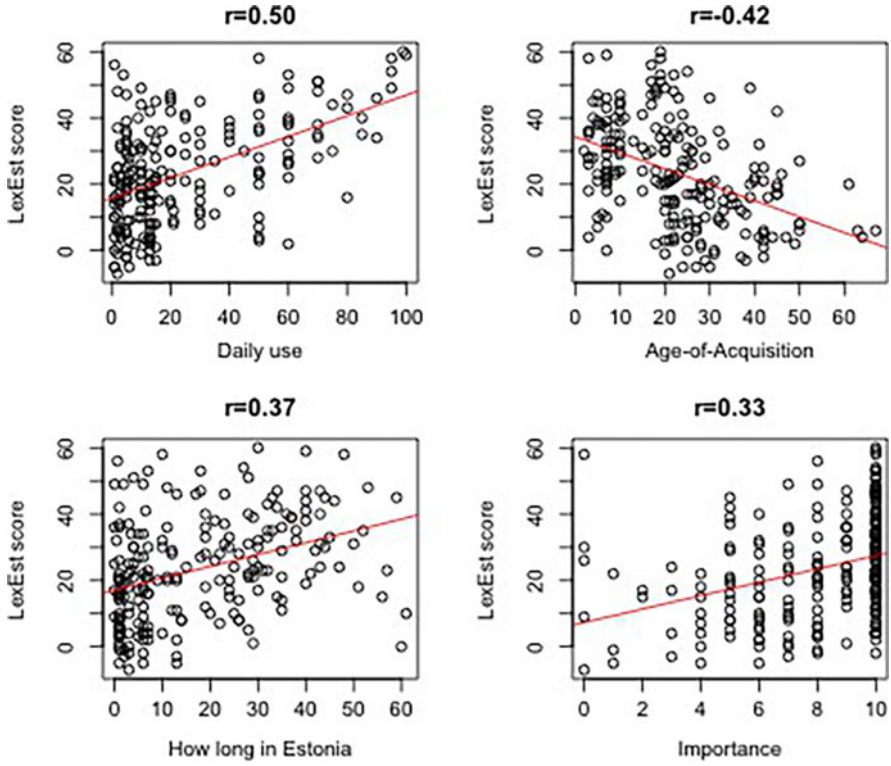


Figure 5. Correlations between LexEst scores and vocabulary knowledge for L2 speakers.

and with C1 level ($Z = 8.59$, $p < .0001$). Participants from the B level course were significantly slower than C1 level ($Z = 5.88$, $p < = .0001$).

Unlike LexEst scores, age of acquisition, duration of stay in Estonia, and importance of learning Estonian were not significant participant variables in the response time analysis.

Finally, the validity of LexEst was evaluated by examining the influence of word frequency on accuracy rates and response latencies, as this factor has consistently shown significant effects in reading and visual lexical decision tasks across multiple studies (Brysbaert et al., 2016; Monsell et al., 1989; Preston, 1935; Rayner, 1998). In line with these studies, our analyses showed clear frequency effects with words of higher frequency coming with shorter response times overall (Spearman $r = -.11$, $p < .0001$), and for L2 speakers separately ($r = -0.15$, $p < .0001$). Moreover, higher frequency words come with higher accuracy rates both overall ($r = -.13$, $p < .0001$) and in particular for L2-s ($r = -.23$, $p < .0001$).

In summary, the response time analysis showed results similar to those of the LexEst score analysis. More advanced L2 speakers not only scored higher on the test but also responded faster. We showed that the higher the participants self-assessed their language proficiency and the more they used Estonian daily, the faster they were. Interestingly, response times were significantly faster at the A2 level compared to the A1 level, while there was no significant difference in LexEst scores between

https://doi.org/10.1017/S0142716425100301 Published online by Cambridge University Press

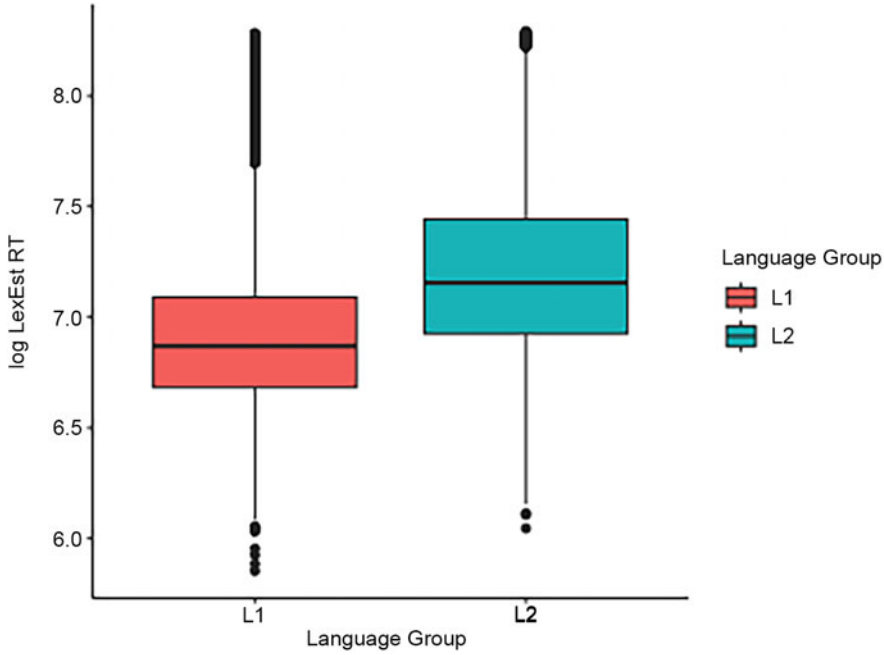


Figure 6. Distribution of LexEst response times in the L1 and L2 groups.

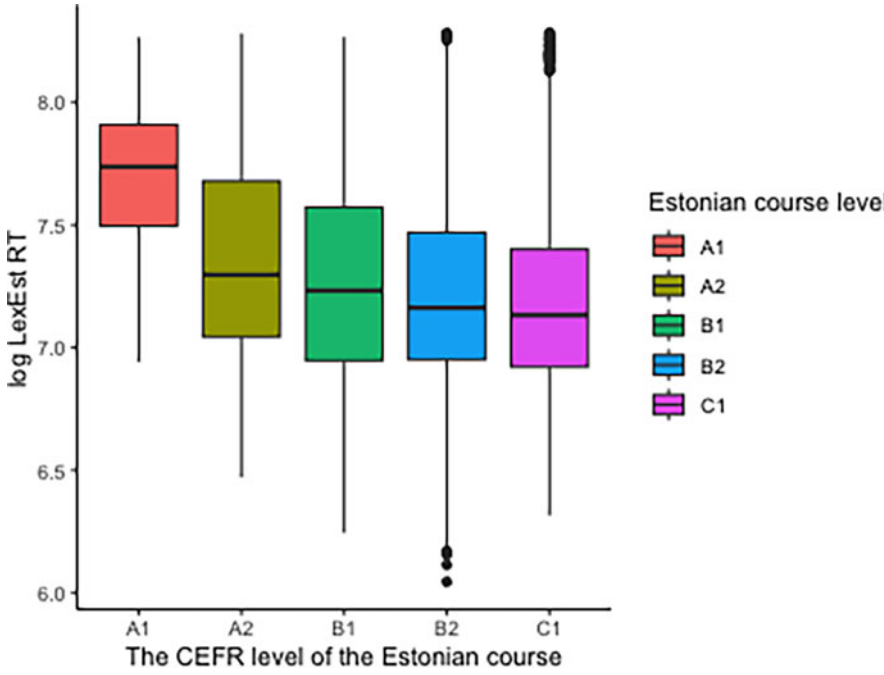


Figure 7. Correlations between LexEst response times and the CEFR level of the Estonian course.

https://doi.org/10.1017/S0142716425100301 Published online by Cambridge University Press

these levels. In addition, we showed that response times and accuracy reflect frequency effects in general, and in particular for L2 speakers.

Discussion and conclusions

The current study introduces the development of LexEst, a freely available vocabulary test for L2 speakers of Estonian. The test is intended to measure vocabulary knowledge among L2 Estonian speakers across various proficiency levels and, by extension, to serve as an indicator of overall L2 Estonian language proficiency. The test is inspired by the original Lexical Test for Advanced Learners of English (LexTALE, Lemhöfer & Broersma 2012) and other more recent LexTALE variants (cf. Amenta et al. 2021; Brysbaert 2013; Izura et al. 2014; Salmela et al. 2021; Zhou & Li 2022).

Our analyses showed higher reliability for the final set of items in the LexEst test than in the English test, returning a Cronbach's alpha of .96, against .81 for the English LexTALE. Also, there was a clear difference in LexEst scores between L1 and L2 speakers. We found that LexEst vocabulary scores correlate well with several indicators of L2 Estonian language proficiency. More specifically, L2 speakers with higher LexEst scores evaluated their own language skills as higher. Additionally, they had participated in an Estonian language course with a higher CEFR level, had started learning Estonian earlier, had resided in Estonia for a longer period of time, considered learning Estonian more important, and used Estonian more frequently on a daily basis. These results are in line with other Lextale-type vocabulary tests, which have indicated a strong relationship between test scores and language proficiency indicators, such as CEFR-language levels and self-assessments, as well as variables closely linked to language proficiency, such as age of acquisition and daily language use (Amenta et al., 2021; Salmela et al., 2021; Zhou & Li, 2022).

With respect to self-assessed language skills, we observed strong and consistent correlations with LexEst scores for all four language modalities. Unexpectedly, the speaking and listening ability ratings were correlated in a similar way with LexEst scores as the reading and writing ability. This result deviates from tests in other languages, which have shown that writing and reading are the abilities that are most closely related to vocabulary knowledge (Milton, 2010; Stæhr, 2008). It is possible that in L2 Estonian teaching, speaking and listening are emphasized to a similar extent as reading and writing, potentially more than in L2 teaching contexts in other countries. However, further research is needed to explore this in more detail as previous work has mainly focused on L2 English.

In previous Lextale-like studies, response times were typically not used for assessing vocabulary knowledge. Here, however, we explored whether and to what extent response times can contribute to gaining insights into vocabulary knowledge skills of L2 speakers. We found that the results for LexEst response times are largely in line with the findings from the score analysis. Specifically, response times were considerably longer for L2 speakers than L1 speakers, and there was a solid frequency effect with high-frequency words eliciting shorter response times and more accurate responses than low-frequency ones (cf. e.g., Balota et al. 2007 in English, Keuleers et al. 2010 in Dutch, Soveri et al. 2007 in Finnish, and Lõo et al. 2018 in Estonian). Moreover, in line with test score results, self-assessed proficiency

https://doi.org/10.1017/S0142716425100301 Published online by Cambridge University Press

ratings correlated positively with LexEst response times for the four modalities. However, the correlations with test scores were considerably stronger than those with response times, indicating that test scores are a more reliable measure of proficiency than response times. Nevertheless, response times can still be considered a potentially valuable variable for assessing vocabulary knowledge.

With respect to the relationship between response times and CEFR course level, it is important to note that response times do not clearly decrease with increasing skill level. In fact, from the B1 level onward, the response times hardly get any faster. There is a slight difference between the B levels and the C1 level, but there is no significant difference between the two B levels. This could indicate that the level of automaticity and processing speed does not change substantially when moving from the B1 to C1 level. As there was a clear increase in LexEst score in the transition from B to C1, one could speculate that the main progression L2 speakers make in this phase is in expanding their vocabulary rather than improving lexical processing speed. However, interestingly, response times on the LexEst task exhibited a clear differentiation between lower-level language course levels (A1 vs. A2), unlike LexEst scores. This pattern could indicate that initial progress in vocabulary, perhaps driven by early language courses, is more focused on developing automaticity than on broadening vocabulary. The CEFR results, at any rate, indicate that it may be worthwhile for future research to explore the role of response times in Lextale-like studies and further investigate whether they could serve as a complementary measure alongside vocabulary test scores. Specifically, it would be interesting to further investigate whether response times are a more sensitive indicator of incremental language processing improvements in the early stages of L2 learning.

The lack of correlations between LexEst response times and language background variables such as age of acquisition, duration of stay in Estonia, and the importance of learning Estonian may reflect the multifaceted nature of lexical processing, especially in an online testing context. Variability in device responsiveness and participants' testing environments may have introduced additional noise into the response time data, potentially obscuring more subtle relationships with individual language history variables.

There are a few methodological choices in the current study that need further discussion. First, we initially started with 135 items and excluded 45 based on IRT analyses and other criteria. We then assessed the reliability and validity of the remaining 90 items, evaluating the validity against specific criteria. However, unlike other studies, we did not conduct a second test round with only the final 90 items. Some studies have done exactly the same (e.g., Brysbaert 2013; Salmela et al. 2021), others have chosen a two-step procedure in which the initial selection of items in a first round was reduced on the basis of IRT analyses, and a second test round was conducted with the reduced item set with new L2 participants (Izura et al., 2014; Lemhöfer & Broersma, 2012; Zhou & Li, 2022). Alongside Brysbaert (2013), we believe that the excluded items are unlikely to have a substantial impact on the accuracy rates of the selected items in the initial test round. Both Izura et al. (2014) and Zhou & Li (2022), with the latter providing more extensive evidence, demonstrated that this is at least true for the Spanish and Portuguese tests. For instance, in the study of Zhou & Li (2022), participants' performance in the first test round closely matched their performance in the second. In both rounds, accuracy

https://doi.org/10.1017/S0142716425100301 Published online by Cambridge University Press

was significantly higher for L1 speakers than for L2 learners, with effect sizes being nearly identical. Additionally, the L2 participants' test scores correlated in a similar way with their self-assessment scores and years spent learning Portuguese in both rounds. We leave it to future research to further investigate this issue in Estonian.

Second, the validation of vocabulary tests often involves comparison with established criterion measures. The original LexTALE test of Lemhöfer & Broersma (2012), for instance, included correlations with a translation test, another lexical decision task and a Quick Placement Task. In the case of LexEst, we do not use these tasks, but we use self-ratings (which correlate with all the other tasks in Lemhöfer & Broersma, 2012) nativity (L1 vs L2), CEFR course level, established subject-variables such as AoA and daily language use, and established item-variables such as word length and word frequency. With all these variables, our LexEst scores show strong correlations (this also holds for response times, but to a lesser extent). These indicators, also used in the validation of earlier LexTALE versions, provide strong evidence that LexEst is a valid measure of Estonian vocabulary knowledge. However, even though we believe that we have thoroughly evaluated the reliability and validity of our vocabulary test, future research could also further extend the validation process by incorporating other lexical tasks, such as translation tasks (see e.g., Lemhöfer & Broersma 2012) and other objective assessments of Estonian language proficiency, e.g., text reading ability (see e.g., Lee et al. 2024).

Another interesting follow-up would be to investigate how the native language of L2 speakers influences their test performance. Here, we choose a global approach, selecting participants from different L1 backgrounds. Although a global approach that can be used for L2 speakers from diverse language backgrounds appears more sustainable, it is important to recognize that language background may play an important role. Lemhöfer & Broersma (2012) found that Dutch participants performed better in the English LexTALE than Korean L2 speakers of English. Puig-Mayenco et al. (2023) found in turn that English LexTALE scores are much more strongly related to scores of a global proficiency measure like the Quick Placement Test for Dutch speakers than for Chinese and Spanish speakers. Also here, it is reasonable to assume that L2 speakers' Estonian vocabulary scores depend on how closely the L1 is related to the Estonian language. Hence, it would be interesting to examine whether participants score better in the LexEst test when their native language is similar to Estonian versus when it is not, for example, comparing Finnish (another Finno-Ugric language) with Slavic Russian L2 speakers.

A final way to extend the current study is to explore whether the LexEst test can also be used for L1 speakers by selecting a more diverse group of native speakers. In general, previous research has shown that L1 speakers perform close to the ceiling level at this test. However, the vocabulary study in Finnish (see e.g., Salmela et al. 2021) showed that vocabulary scores can depend on education level and the final exam grade in the lower secondary school. In the current study, the majority of our L1 participants had at least an undergraduate degree, so the variability in educational background was limited. Another way to include a more diverse L1 population would be to test L1 children with LexEst.

In summary, LexEst is a quick 5-minute vocabulary test for Estonian as a second language, freely available at https://lexest.ut.ee/ and as a paper-and-pencil test to be found in the Appendix C and D. Similarly to previous studies, LexEst returns clear

https://doi.org/10.1017/S0142716425100301 Published online by Cambridge University Press

differences between L1 and L2 speakers, and LexEst scores correlate well with self-assessment ratings, CEFR proficiency level, exposure to Estonian, age of acquisition, and motivation to learn Estonian. This suggests that LexEst provides an objective and attractive alternative for self-assessments and language background questionnaires to measure Estonian language proficiency mostly in research settings.

Replication package. The data and the code for the item assessment are available in: https://osf.io/y42xv/.

Acknowledgments. We are grateful for everyone who participated in our study and for Tomi Rautaoja for all his help with the test development and data collection. This research was funded by the Kadri, Nikolai, and Gerda Rõuk research fund, as well as by the Estonian Research Council grant PSG743.

Notes

1 https://rahvaloendus.ee/en/news/population-census-76-estonias-population-speak-foreign-language (07.12.2024).
2 The Common European Framework of Reference for Languages.
3 https://web.meis.ee/testest/ (07.12.2024).
4 https://sonaveeb.ee/teacher-tools (07.12.2024).
5 We considered as foreign origin words lexical items that are probably perceived as a loanword from other major languages by a non-linguist, for example *bus* is *buss* in Estonian.
6 https://ville.utu.fi/
7 https://sonaveeb.ee/teacher-tools (07.12.2024).

References

Allen, P. A., Madden, D. J., & Crozier, L. C. (1991). Adult age differences in letter-level and word-level processing. *Psychology and Aging*, **6**(2), 261.

Alzahrani, A. (2024). LexArabic: a receptive vocabulary size test to estimate Arabic proficiency. *Behavior Research Methods*, **56**(6), 5529–5556.

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.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The english lexicon project. *Behavior Research Methods*, **39**, 445–459.

Bertram, R., Rautaoja, T., Holopainen, S., Häikiö, T., Enges, P., Hyönä, J., Lehtonen, M., Pugh, K. R., Rueckl, J. G., Salmela, R., Siegelman, N., & Räsänen, P. (2025). Assessing vocabulary skills of school children aged 9 to 15 in finland: tracking the gender and home language gap. Preprint https://doi.org/10.21203/rs.3.rs-6448049/v1.

Brysbaert, M. (2013). Lextale FR a fast, free, and efficient test to measure language proficiency in french. *Psychologica Belgica*, **53**(1), 23–37.

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: evidence from the Dutch Lexicon Project 2. *Journal of Experimental Psychology: Human Perception and Performance*, **42**(3), 441–458.

Chan, I. L. & Chang, C. B. (2018). LEXTALE CH: A quick, character-based proficiency test for mandarin chinese. In *Proceedings of the Annual Boston University Conference on Language Development*: Cascadilla Press.

Crocker, L. & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. ERIC.

Dijkgraaf, A., Hartsuiker, R. J., & Duyck, W. (2017). Predicting upcoming information in native-language and non-native-language auditory word recognition. *Bilingualism: Language and Cognition*, **20**(5), 917–930.

Dinno, A. (2024). *dunn.test: Dunn's Test of Multiple Comparisons Using Rank Sums*. R package version 1.3.6.

https://doi.org/10.1017/S0142716425100301 Published online by Cambridge University Press

Grosjean, F. (1998). Studying bilinguals: methodological and conceptual issues. *Bilingualism: Language and Cognition*, **1**(2), 131–149.

Harrington, M. (2006). The lexical decision task as a measure of L2 lexical proficiency. *EUROSLA yearbook*, **6**(1), 147–168.

Izura, C., Cuetos, F., & Brysbaert, M. (2014). Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicologica: International Journal of Methodology and Experimental Psychology*, **35**(1), 49–66.

Kaalep, H.-J. & Muischnek, K. (2005). The corpora of Estonian at the University of Tartu: the current situation. In *Proceedings of the Second Baltic Conference on Human Language Technologies* (pp. 267–272).

Kallas, J., Koppel, K., Pool, R., Tsepelina, K., Üksik, T., Alp, P., & Epner, A. (2021). Eesti keele kui teise keele õpetaja tööriistad Eesti Keele Instituudi keeleportaalis Sõnaveeb [Tools for teaching Estonian as a second language in the language portal Sõnaveeb of the Institute of the Estonian Language]. *Eesti Rakenduslingvistika Ühingu Aastaraamat*, **17**, 61–80.

Kaushanskaya, M., Blumenfeld, H. K., & Marian, V. (2019). The language experience and proficiency questionnaire (LEAP-Q): ten years later. *Bilingualism: Language and Cognition*, **23**(5), 945–950.

Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice effects in large-scale visual word recognition studies: a lexical decision study on 14,000 dutch mono- and disyllabic words and nonwords. *Frontiers in Psychology*, **1**, 174. https://doi.org/10.3389/fpsyg.2010.00174

Laufer, B. & Nation, P. (2001). Passive vocabulary size and speed of meaning recognition: Are they related? *EUROSLA yearbook*, **1**(1), 7–28.

Laufer, B. & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, **22**(1), 15–30.

Lee, S. T., van Heuven, W. J., Price, J. M., & Leong, C. X. R. (2024). LexMAL: a quick and reliable lexical test for Malay speakers. *Behavior Research Methods*, **56**(5), 4563–4581.

Lemhöfer, K. & Broersma, M. (2012). Introducing LexTALE: a quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, **44**(2), 325–343.

Lippus, P., Aare, K., Malmi, A., Tuisk, T., & Teras, P. (2021). Phonetic Corpus of Estonian Spontaneous Speech v1.2.

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.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, **50**(4), 940–967.

Meara, P. & Jones, G. (1988). *Vocabulary Size as a Placement Indicator*. ERIC.

Milton, J. (2010). The development of vocabulary breadth across the cefr levels. *Communicative Proficiency and Linguistic Development: Intersections between SLA and Language Testing Research*, **1**, 211–232.

Miralpeix, I. & Meara, P. (2014). Knowledge of the written word. In *Dimensions of Vocabulary Knowledge* (pp. 30–44). London: Macmillan Education UK.

Miralpeix, I. & Muñoz, C. (2018). Receptive vocabulary size and its relationship to efl language skills. *International Review of Applied Linguistics in Language Teaching*, **56**(1), 1–24.

Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: where are they? *Journal of Experimental Psychology: General*, **118**(1), 43–71.

Nation, P. (1983). Testing and teaching vocabulary. *Guidelines*, **5**, 12–15.

Preston, K. A. (1935). The speed of word perception and its relation to reading ability. *The Journal of General Psychology*, **13**(1), 199–203.

Puig-Mayenco, E., Chaouch-Orozco, A., Liu, H., & Mart´ın-Villena, F. (2023). The lextale as a measure of l2 global proficiency: A cautionary tale based on a partial replication of lemh¨ofer and broersma (2012). *Linguistic Approaches to Bilingualism*, **13**(3), 299–314.

R Development Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, **124**(3), 372.

Read, J. (2000). *Assessing Vocabulary*. Cambridge University Press.

Rizopoulos, D. (2007). ltm: an R package for latent variable modeling and item response analysis. *Journal of Statistical Software*, **17**, 1–25.

https://doi.org/10.1017/S0142716425100301 Published online by Cambridge University Press

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.

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, **25**, 289–310.

Soveri, A., Lehtonen, M., & Laine, M. J. (2007). Word frequency and morphological processing in finnish revisited. *The Mental Lexicon*, **2**(3), 359–385.

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, **36**(2), 139–152.

Üksik, T., Kallas, J., Koppel, K., Tsepelina, K., & Pool, R. (2021). Estonian as a second language teacher's tools. In *Proceedings of the 16th Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 130–134).

Vaks, A., Padrik, M., & Vihman, V. (2025). Eestikeelse sõnavaratesti väljatöötamine kakskeelsetele lastele [The development of an Estonian vocabulary test for bilingual children]. *Eesti Rakenduslingvistika Uhingu Aastaraamat*, **21**, 327–343.

Wen, Y., Qiu, Y., Leong, C. X. R., & van Heuven, W. J. (2024). LexCHI: A quick lexical test for estimating language proficiency in Chinese. *Behavior Research Methods*, **56**(3), 2333–2352.

Wilson, L. T. (1953). "cloze procedure": a new tool for measuring readability. *Journalism Quarterly*, **30**(4), 415–433.

Zareva, A., Schwanenflugel, P., & Nikolova, Y. (2005). Relationship between lexical competence and language proficiency: variable sensitivity. *Studies in Second Language Acquisition*, **27**(4), 567–595.

Zell, E. & Krizan, Z. (2014). Do people have insight into their abilities? A metasynthesis. *Perspectives on Psychological Science*, **9**(2), 111–125.

Zhang, S. & Zhang, X. (2022). The relationship between vocabulary knowledge and L2 reading/listening comprehension: a meta-analysis. *Language Teaching Research*, **26**(4), 696–725.

Zhou, C. & Li, X. (2022). LextPT: a reliable and efficient vocabulary size test for L2 Portuguese proficiency. *Behavior Research Methods*, **54**(6), 2625–2639.

Zimmerman, K. J. (2004). *The Role of Vocabulary Size in Assessing Second Language Vocabulary*. Brigham Young University.

https://doi.org/10.1017/S0142716425100301 Published online by Cambridge University Press

Appendix A: Final LexEst item list with English translations

Words (60): pilk 'to look', kütma 'to heat', mälu 'memory', nali 'joke', tõbi 'disease', ere 'bright', kaine 'sober', katk 'plague', kimp 'bouguet', kraav 'ditch', lärm 'noise', lõtv 'flaccid', madal 'low', nukker 'sad', pala 'piece', pruukima 'to use', putukas 'insect', siiras 'sincere', sõimama 'to curse', suusk 'ski', taat 'old man', täpp 'dot', toores 'raw', helge 'bright', jõhker 'brutal', kõha 'couch', kuduma 'to knit', muhk 'bump', naps 'drink', pekk 'lard', piits 'whip', puri 'sail', rase 'pregnant', roomama 'to crawl', sõõm 'sip', tuisk 'snowstorm', ümar 'circular', vaen 'feud', võigas 'gruesome', äpu 'helpless', kobe 'decent', kõnts 'filth', kriim 'scratch', laitma 'to blame', lauge 'gentle (slope)', niru 'lousy', räige 'vulgar', riisuma 'to rake', sagima 'to bustle about', sahver 'pantry', seedima 'to digest', sirp 'sickle', tahm 'soot', tärn 'asterisk', tikkima 'to em- broider', tragi 'gumptious', adru 'seaweed', käbe 'quick', klii 'bran', susima 'to gossip'.

Pseudowords (30): träll, trahn, nibe, karal edima, jabu, leegas, pasuma, ölbas, mura, pammas, mook, rainama, tägar, pläva, köim, megel, teep, käver, riits, sabul, laberin, kora, vunuma, örv, mitsi, jast, pübama, kösima, niim.

Appendix B: Native languages of LexEst L2 speakers

Table 3. Distribution of native languages of L2 speakers

Language	Number of L2 speakers
Russian	110
German	26
Finnish	17
English	11
Ukrainian	8
Latvian	4
Hungarian	4
Swedish	4
French	3
Polish	3
Dutch	3
Turkish	2
Spanish	2
Portuguese	2
Italian	2
Chinese	2
Tatar	1
Norwegian	1
Nepali	1
Komi	1

(*Continued*)

https://doi.org/10.1017/S0142716425100301 Published online by Cambridge University Press

Table 3. (*Continued*)

Language	Number of L2 speakers
Hausa	1
Bulgarian	1
Azerbaijani	1
Arabic	1

Appendix C: LexEst vocabulary test in Estonian

Jägenevas nimekirjas on päris ja väljamõeldud eesti keele sõnad. Sinu ülesanne on otsustada, millised neist on päriselt eesti keeles olemas. Tee linnuke nende sõnade ette, mis sinu arvates on päris eesti keele sõnad. Lahenda testi ülevalt alla tulpade kaupa

Kui oled kindel, et tegemist on eesti keele sõnaga, aga sa ei tea sõna täpset tähendust, siis tee selle sõna ette ikkagi linnuke. Kui sa pole kindel, kas see s˜ona on eesti keeles olemas, ära selle ette linnukest tee. Valede vastuste eest saab miinuspunkte. Test ei ole aja peale. Tee testi iseseisvalt, ilma teiste ja sõnaraamatu abita.

Millised on päris eesti keele sõnad?

megel	kõsima	lõtv	trahn	siiras
õrv	teep	kriim	vununa	nibe
tikkima	kimp	pläva	pasuma	kuduma
pruukima	sõõm	pübama	tõbi	täpp
riisuma	ümar	piits	tärn	riits
mura	tahm	tägar	klii	kraav
helge	rase	laitma	mitsi	tragi
adru	äpu	edima	sirp	niru
lauge	kaine	toores	nali	susima
kütma	sahver	lärm	leegas	pekk
räige	nukker	võigas	sabul	jast
ere	jabu	köha	roomama	niim
suusk	kobe	taat	putukas	rainama
madal	jõhker	kõim	ölbas	sagima
seedima	puri	karal	träll	käver
pala	muhk	tuisk	pilk	pammas
kõnts	mook	katk	sõimama	mälu
kora	naps	laberin	käbe	vaen

https://doi.org/10.1017/S0142716425100301 Published online by Cambridge University Press

Appendix D: LexEst vocabulary test in English

The following is a list of real and made-up Estonian words. Your task is to decide which of them actually exist in the Estonian language. Select the ones that you believe are Estonian words. Solve the task column by column from top to bottom, starting from the left.

If you are certain that a word is an Estonian word but do not know its exact meaning, still select that word. If you are unsure whether a word exists in Estonian, do not select it. You will lose points for incorrect answers. The test is not timed. Complete the test on your own, without help from others or a dictionary.

Which ones are real Estonian words?

megel	kõsima	lõtv	trahn	siiras
õrv	teep	kriim	vununa	nibe
tikkima	kimp	pläva	pasuma	kuduma
pruukima	sõõm	pübama	tõbi	täpp
riisuma	ümar	piits	tärn	riits
mura	tahm	tägar	klii	kraav
helge	rase	laitma	mitsi	tragi
adru	äpu	edima	sirp	niru
lauge	kaine	toores	nali	susima
kütma	sahver	lärm	leegas	pekk
räige	nukker	võigas	sabul	jast
ere	jabu	köha	roomama	niim
suusk	kobe	taat	putukas	rainama
madal	jõhker	kõim	õlbas	sagima
seedima	puri	karal	tr¨all	käver
pala	muhk	tuisk	pilk	pammas
kõnts	mook	katk	sõimama	mälu
kora	naps	laberin	käbe	vaen

Cite this article: Lõo, K., Leppik, K., Malmi, A., Bleive, A., & Bertram, R. (2025). Introducing LexEst: a quick and efficient vocabulary test for assessing vocabulary knowledge in L2 Estonian. *Applied Psycholinguistics*. https://doi.org/10.1017/S0142716425100301

https://doi.org/10.1017/S0142716425100301 Published online by Cambridge University Press