Recently, there has been an increased interest in the development of new vocabulary tests, as older tests have been criticized because of their use of the word family as counting unit (Kremmel, 2016), the lack of a test section focusing on the 1,000 most frequent words, and the old frequency lists used (Webb & Sasao, 2013).

The aim of the present study is to compare EFL learners’ performance on three frequency-based vocabulary levels tests: VLT (Schmitt et al., 2001), the updated VLT (uVLT) (Webb et al., 2017), and the VocabLab-test (Peters et al., 2019) to determine whether the tests correlate and result in comparable test scores per total and frequency band. The tests differ in (1) the frequency lists used to develop the tests, (2) the counting unit (lemma vs. word family), (3) division into test sections, (4) inclusion of cognates, (5) the inclusion of an “I don’t know”-option, and (6) test format (multiple choice vs. matching).

Data were collected from 94 Dutch-speaking learners, 172 Vietnamese-speaking learners, and 191 Spanish-speaking learners taking the aforementioned tests. The analyses showed that the three tests were highly correlated (rho = .58-.92). The correlations for the test sections corresponding to different frequency bands ranged from .52 to .88. The highest correlations were found between the VLT and the VocabLab-test, the lowest between the VLT and the uVLT. Second, related samples Wilcoxon tests showed that there were differences between learners’ test performance in the total scores as well as in the scores per level. Finally, an analysis of the test sections showed that the cutting score of 26 or 27 out of 30 per level, which has often been proposed as a criterion for mastery, resulted in considerable differences between the three tests.