openjournals.nl

Dutch Journal of Applied Linguistics
https://www.dujal.nl | E-ISSN 2211-7253

Academic vocabulary size as a predictor for early academic achievement in higher education: A case study of an easy-to-administer, low-stakes, post-entry test

Pieterjan Bonne[1,2] and Jordi Casteleyn[2]

[1] *Artevelde University of Applied Sciences* | [2] *University of Antwerp*

Abstract This study investigates the predictive value of an academic vocabulary size screening test in Dutch for early academic achievement in higher education, in a context where Dutch is the predominant L1 of instruction. The research compares the performance of an academic vocabulary size screening test with a general vocabulary counterpart, revealing that the academic test exhibits slightly stronger reliability and similar associations with personal and educational characteristics, often with more pronounced effect sizes. Regression analyses demonstrate that academic vocabulary size significantly contributes to predicting Grade Point Average (GPA) and Cumulative Study Efficiency (CSE). However, prior secondary education remains the most critical predictor of academic achievement. The findings suggest that an academic vocabulary size screening test can be a valuable tool for supporting the initial success of incoming higher education students.

Keywords vocabulary knowledge, vocabulary size, academic vocabulary, academic achievement, higher education, student success

Samenvatting Deze studie onderzoekt de voorspellende waarde van een screeningstest academische woordenschatgrootte Nederlands voor studiesucces in het eerste semester in het hoger onderwijs met Nederlands als dominante instructietaal. Het onderzoek vergelijkt de prestaties van een screeningstest voor academische woordenschatgrootte met die van een screeningstest voor algemene woordenschatgrootte, en toont aan dat de academische test iets sterker scoort op betrouwbaarheid en vergelijkbare verbanden vertoont met persoonlijke en onderwijskundige kenmerken, vaak met meer uitgesproken effectgroottes. Regressieanalyses laten zien dat academische woordenschatgrootte significant bijdraagt aan het voorspellen van gemiddelde studieresultaten (GPA) en behaalde studiepunten (CSE). Toch blijft het eerder genoten secundair of voortgezet onderwijs de belangrijkste voorspeller van academisch succes. De bevindingen suggereren dat een screeningstest voor academische woordenschatgrootte een waardevol instrument kan zijn om het aanvankelijke succes van nieuwe studenten in het hoger onderwijs te ondersteunen.

Trefwoorden woordenschat, woordenschatgrootte, academische woordenschat, studiesucces, hoger onderwijs

Article history
Received: October 29, 2024
Accepted: July 2, 2025
Online: November 24, 2025

Corresponding author
Pieterjan Bonne,
pieterjan.bonne@arteveldehs.be

Author contributions
Funding information
Statement of interest
Statement of technology use
Supporting information
See p. 20–21

How to cite Bonne, P. and Casteleyn, J. (2025). 'Academic vocabulary size as a predictor for early academic achievement in higher education: A case study of an easy-to-administer, low-stakes, post-entry test', *Dutch Journal of Applied Linguistics* 14: DOI 10.51751/dujal19780

1 Introduction

Language proficiency significantly impacts academic achievement in higher education for both L1 and L2 students (e.g. Feng et al., 2023; Heeren, 2024). Specifically, vocabulary as a key facilitator of the four language skills (Alderson, 2005) is suggested to play a pivotal role in explaining variations in language performance (Qian & Lin, 2023) and, consequently, in academic achievement (Irvin & Blankenship, 2022; Trenkic & Warmington, 2019; Warnby, 2023). Within higher education, general academic and specialist subject-specific vocabulary constitutes an essential element of academic discourse (e.g. Hyland, 2006). Although often acquired within the context of higher education itself (Bonne & Casteleyn, 2021), this specific vocabulary is used in lectures and reading material from day one. Knowledge of academic vocabulary can significantly and measurably enhance academic success, beyond the impact of general vocabulary knowledge alone, especially in the context of more proficient language users like students in higher education (Masrai & Milton, 2021).

Hence, a screening test of academic receptive vocabulary holds promise as a valuable screening tool for incoming students. This study contributes to the literature by introducing a newly developed, easy-to-administer Dutch academic vocabulary size screening test and by examining its predictive value specifically in a higher education context where the language of instruction is also the students' predominant L1. Furthermore, emphasizing academic vocabulary, rather than general or subject-specific vocabulary, increases the relevance for test-takers, the usefulness for staff to engage in a dialogue on academic language skills and the practicality for institutions.

2 Theoretical background

2.1 Academic achievement

A commonly used measure of academic achievement is Grade Point Average (GPA), the average of scores across all enrolled courses. GPA is often employed as a proxy for successful academic communication (Masrai & Milton, 2021). An alternative measure is Cumulative Study Efficiency (CSE). CSE assesses the ratio of acquired credits to attempted credits. CSE is often used to determine a student's eligibility for continued study (Heeren et al., 2021). In this respect, CSE emphasizes progress, whereas GPA rather measures success.

2.2 Vocabulary knowledge

Vocabulary knowledge has been linked to academic achievement as a robust indicator of overall academic literacy (Milton, 2013; Read, 2015) and positively impacts the transition

to higher education and a student's academic achievement (Daller et al., 2021; Fonteyne et al., 2017; Heeren et al., 2021; Milton & Treffers-Daller, 2013; Paribakht & Webb, 2016; Vandervieren & Casteleyn, 2020).

Higher education complements high-frequent words with two sets of specific vocabulary that are not part of ordinary language: academic vocabulary – words prevalent in successful academic communication, and specialist vocabulary – subject-specific words (Coxhead, 2000). Mastering these vocabulary sets is essential for students to cope with educational tasks (Coxhead, 2016) and to acquire content (Peters, 2017).

Assuming all incoming students possess adequate vocabulary knowledge would be unwise. Academic and subject-specific vocabulary, as clear examples of extended language cognition (Hulstijn, 2015), lies beyond the standard lexicon of all L1 speakers, including students entering higher education. Demographic shifts (Deygers et al., 2017; Wingate, 2015) further compound this challenge, particularly for students from low-literacy backgrounds, multilingual backgrounds, or those with prior education lacking a focus on continuing education (Elder, 2017; Geldof, 2014; Glorieux et al., 2014; van Kalsbeek & Kuiken, 2014). Additionally, academic discourse itself poses a challenge. Specific vocabulary constitutes a substantial proportion of academic texts (Warnby, 2023): academic words account for approximately 8-10 % of texts (Nation, 2013) and subject-specific vocabulary up to 5 % (Hyland & Tse, 2007). With such a lexical coverage, the texts inherently challenge students who require knowledge of 95 % or 98 % of the words for general or detailed understanding (Laufer & Ravenhorst-Kalovski, 2010; Van Zeeland & Schmitt, 2013).

2.3 Vocabulary screening tests

Considering the impact of vocabulary knowledge on academic achievement, vocabulary screening tests are useful tools. These tests can be differentiated on four aspects. A first differential is the language tested. Vocabulary screening tests are used to assess vocabulary knowledge of native languages (L1) (Heeren et al., 2021; Irvin & Blankenship, 2022; Vandervieren & Casteleyn, 2020; Welie et al., 2021) and non-native languages (L2), especially in the context of English for academic purposes (Gu & So, 2015; Masrai & Milton, 2018; Skjelde & Coxhead, 2020; Warnby, 2023). A second differential is the selected vocabulary for the test. Both general vocabulary (Irvin & Blankenship, 2022; Vandervieren & Casteleyn, 2020) and academic vocabulary (Heeren et al., 2021; Warnby, 2023) are tested in relation to academic achievement, sometimes together (Masrai & Milton, 2018). Vocabulary screening tests only exceptionally include subject-specific vocabulary (Masrai et al., 2021). A third differential is whether the screening tests measure vocabulary size (Heeren et al., 2021; Masrai & Milton, 2021; Milton & Treffers-Daller, 2013), depth (Alsahafi, 2023), or both (Lemmouh, 2010; Scheepers, 2019; Welie et al., 2021). A final differential is whether the screening tests look at receptive or productive vocabulary knowledge (Paribakht & Webb, 2016). Receptive tests tend to have a high

Table 1 Overview of vocabulary screenings used in relation to academic achievement

Name	Reference	Language	Academic/ general	Receptive/ productive	Size/ depth	Time in minutes
The Peabody Picture Vocabulary Test (PPVT)	Dunn & Dunn, 2007	English	General	Receptive	Size	10 to 15
The Vocabulary Levels Test (VLT)	Nation, 1983; Schmitt et al., 2001	English	General	Receptive	Size	30 to 45
The Vocabulary Levels Test – Academic expension (VLT-Ac)	Nation, 1983; Schmitt et al., 2001	English	Academic	Receptive	Size	30 to 45
The Productive Vocabulary Levels Test (PVLT)	Laufer & Nation, 1999	English	General	Productive	Size	30 to 45
Academic Vocabulary Test (AVT)	Pecorari et al., 2019	English	Academic	Receptive	Size	15 to 20
General Vocabulary Knowledge Test (GVT)	Masrai & Milton, 2021	English	General	Receptive	Size	10
Academic Vocabulary Size Test (AVST)	Masrai & Milton, 2018	English	Academic	Receptive	Size	10
XK_Lex	Al-Masrai, 2009 Keuleers et al., 2015	English Dutch	General	Receptive	Size	5 to 10
The Academic Reading Vocabulary screening (ARV)	De Wachter & Heeren, 2013	Dutch	Academic	Receptive	Depth	30
Academische Woordenschat Test (AWT)	Welie et al., 2021	Dutch	Academic	Receptive	Size Depth	40

degree of reliability and validity. Productive tests are more contested and do not appear to show stronger relationships than receptive tests (Lemmouh, 2010).

A substantial variety exists in vocabulary screening tests used in relation to academic achievement, as shown in Table 1.

For the present study, three requirements were identified for a vocabulary screening test focusing on academic achievement of incoming students. Firstly, the screening test should link to receptive skills. Students are first exposed to academic vocabulary in lectures and reading material. Vocabulary knowledge also gradually develops from receptive towards productive mastery (Schmitt, 2014). Secondly, the screening test should be easy to administer, considering students' and staff's time. Due to practical issues (Qian & Lin, 2023) and because depth is harder to approach from a theoretical and practical perspective (Schmitt, 2014) there is a preference for vocabulary size. Additionally, size relates closer to receptive knowledge. Thirdly, the screening test should be relevant for higher education. For L1 speakers entering higher education, including academic vocabulary in the language of instruction enhances relevance (Masrai & Milton, 2021).

As far as our knowledge extends, no screening test specifically assesses receptive Dutch academic vocabulary size in the context of academic achievement in higher education.

3 Present study

This study aims to investigate whether an easy-to-administer, low-stakes, post-entry screening test of receptive academic vocabulary size of the instruction language, Dutch as L1, can effectively support the academic achievement of incoming students. In this respect, the following two research questions can be formulated:

- How does an L1 academic vocabulary screening test compare to an L1 general vocabulary screening test in assessing students in higher education?
- To what extent does an L1 academic vocabulary size screening test predict academic achievement of incoming students?

4 Method

4.1 Context and participants

This study took place at a university college in the Dutch-speaking region of Belgium, Flanders. The Flemish higher education system is characterized by relatively open access, with standardized entrance exams in place for only Medicine, Dentistry and Veterinary Medicine. The university college has around 15,000 students and 1,500 staff in 24 bachelor degrees, 5 advanced bachelors, 12 associate degrees and 41 post-graduate degrees in five domains: Education, Health and Care, Business and Management, People and Society, and Communication, Media and Design. The primary language of instruction is Dutch.

Participants were recruited during a support initiative with voluntary participation on listening skills, looking at vocabulary, metacognition and note-taking. The initiative ran during the introductory courses in the third week of September just before the official start of the academic year (n = 172) and during the first four weeks of the academic year (n = 124). The screening test was administered at the introductory course of four academic years: 2018-2019, 2019-2020, 2020-2021, and 2022-2023. Due to Covid restrictions, the initiative did not run in the 2021-2022 academic year. The sessions during the year were discontinued after one academic year: 2018-2019. Although all students were welcome, communication targeted first-year students for the introductory sessions and older years for the sessions in the semester.

The screening test was administered at the beginning of the support initiative. As part of the paper teaching materials, participants received an informed consent, the

Table 2 Division of participants across academic years: time of data collection and experience in higher education

		18-19	19-20	20-21	22-23	Total
Total number of participants per academic year		170	37	30	59	296
Time of data collection	Introductory session	46	37	30	59	172
	Session during the semester	124	0	0	0	124
Experience in higher education	No	76	34	27	55	192
	Yes	94	3	3	4	104

general and academic vocabulary screening tests, and a survey on personal and educational characteristics. All screening test materials were completed during a single testing session.

A subset of the participants (n = 71) voluntarily retook the two vocabulary screening tests after a minimum of one month (33 days) and a maximum of 3 months (71 days) (M = 51.21, SD = 10.6). The retesting was conducted online, following the same instructions as the paper version.

In total, the present study includes 296 participants across four academic years: 2018-2019 (n = 170), 2019-2020 (n = 37), 2020-2021 (n = 30) and 2022-2023 (n = 59). As shown in Table 2, 172 participants participated during the introductory sessions and 124 during the first semester. Of all participants, 192 were incoming students, 104 had experience in higher education.

To determine whether our sample adequately represents the broader student population, we surveyed participants on personal and educational characteristics, enabling comparisons in the performance and functionality of the academic and general vocabulary size screening tests. 276 participants completed the entire survey, the others filled in the survey in varying degrees. Therefore, the number of participants of whom the information is available is mentioned per characteristic. The mean age of the participants (n = 290) was 21 years (M = 20.6, SD = 4.58, min = 17.15, max = 53.19). Of the participants (n = 296), 73% identified as female, 27% as male. Regarding language background (n = 276), 85.9% of the participants reported Dutch as their only native language. 9.1% identified Dutch and another language as their native languages, and 5.1% of participants reported only another language than Dutch as their native language. 24% of participants reported a disability. These disabilities included physical, developmental, behavioral, emotional, and sensory impairments. The most frequently reported disabilities were dyslexia/dysorthography (n = 28) and attention deficit disorder (n = 10).

Table 3 Percentage of participants divided per department

Department	Number of participants	Percentage of participants
Health and Care	90	30.4%
Education	77	26.0%
Business and Management	58	19.6%
People and Society	38	12.8%
Communication, Media and Design	28	9.5%
Total	**291**	**100%**

Table 4 Percentage of participants divided per form of prior education

Form of prior education	Number of participants	Percentage of participants
General Secondary School (ASO)	138	46.6%
Technical Secondary School (TSO)	117	39.5%
Vocational Secondary School (BSO)	24	8.1%
Art Secondary School (KSO)	9	3.0%
Secondary degree from the Netherlands	7	2.3%
Central or state examining board	1	0.3%
Total	**296**	**100%**

A total of 104 participants had experience in higher education, varying from one to twelve semesters (M = 4.73, SD = 2.47). Fourteen participants held a bachelor's degree, two a master's degree. The vast majority of the participants were enrolled in a bachelor's programme (n = 290). One participant followed a graduate programme and three participants were enrolled in a master's programme at a different institution. As shown in Table 3, participants from the own institution (n = 291) came from all five departments of the university college.

The majority of participants completed either General or Technical Secondary School, as shown in Table 4.

Regarding the educational background of the parents (n = 283), 27.6% of the participants had parents without a higher education degree. 40.3% of the participants had two parents with a higher degree, only the father had a degree of 12.7% of the participants and only the mother of 14.8% of the participants. 4.6% were unsure whether their parents had a degree or not.

4.2 Instruments

4.2.1 *General vocabulary size*

To measure general vocabulary size, an XLex checklist test was used (Brysbaert, Stevens, Mandera, & Keuleers, 2014). This test assesses broad word recognition across a wide range of word frequency levels and topics. Its strong correlation with overall language proficiency makes it a reliable indicator of general vocabulary knowledge. The test is unspeeded and typically takes between 5 to 10 minutes to complete.

The test involves a lexical decision task where participants indicate words they know or can use based on 100 sequences of letters. 77 sequences constitute actual words, while 23 are pseudowords included to discourage guessing. The resulting score, expressed as percentage, is calculated as follows: ((number of correctly identified (= checked) actual words/77*100) + (number of correctly identified (= not-checked) false words/23*100)) /2.

To assess internal consistency, Cronbach's alfa was calculated based on 296 participants who had completed the general vocabulary screening test and 99 of the 100 test items. One item with zero variance was removed. The Cronbach's alpha value (α = .88) indicates a good internal consistency among the items. For the real words, Cronbach's alfa was calculated based on 76 of the 77 test items. One item with zero variance was removed. The Cronbach's alpha value (α = .90) again indicates an excellent internal consistency among the items. For the pseudowords, Cronbach's alfa was calculated based on 22 of the 23 test items. One item with zero variance was removed. The Cronbach's alpha value (α = .70) indicates an acceptable internal consistency among the items.

4.2.2 *Academic vocabulary size*

The academic vocabulary size screening test for this research was developed in parallel with the general vocabulary size screening test (Brysbaert et al., 2014). The selection of the 77 real academic words drew on a corpus of 5,875 Dutch academic word families (Bonne & Casteleyn, 2021). This corpus was categorized into four frequency bands based on the number of sources in which at least one item of the word family appeared. From each band, Excel randomly selected 19 items. One extra item was drawn from the band of the highest frequency. The 23 pseudowords were drawn from an existing list (Brysbaert et al., 2014). Similarly, this list was divided into four frequency bands based on correct recognition as pseudowords in Flanders. Excel randomly selected six words from three quarters and five from the quarter with the lowest frequency.

To assess internal consistency, Cronbach's alfa was calculated based on data from 294 participants who had completed the academic vocabulary screening test and 99 of the 100 items. One item with zero variance was removed. The Cronbach's alpha value (α = .91) indicates an excellent internal consistency among the items. For the real words, Cronbach's alfa was computed based on all of the 77 test items and also shows excellent internal consistency (α = .94). For the pseudowords, Cronbach's alfa was calculated based

on 22 of the 23 test items. One item with zero variance was removed. The Cronbach's alpha value (α = .77) indicates an acceptable internal consistency among the items.

4.2.3 *Personal and educational characteristics*

A survey gathered self-reported information on the participants' personal and educational characteristics. It considered characteristics known to influence academic achievement such as gender, language background, the form of followed secondary education and parental education. The survey included 16 questions.

4.2.4 *Academic achievement*

This research focuses on early academic achievement to minimize the impact of spontaneous language acquisition and language support initiatives. Additionally, it prevented the loss of participants who discontinued their education after the first exam period in January. The grades of this exam period were collected from the participants' records. Based on these grades, GPA and CSE were calculated.

All procedures were in accordance with the ethical standards of the institution where the research was carried out. Informed consent was obtained from all individual participants included in the study. All data collection adhered to GDPR regulations and all rules and regulations set up by the university college.

4.3 Method of analysis

To assess whether the different academic years could be combined in the analysis, the differences among groups were calculated for general and academic vocabulary size using ANOVA and Kruskal-Wallis tests. For general vocabulary size, Levene's test indicated equal variances ($F(3,292) = 0.290$, $p = .832$). A one-way ANOVA revealed a significant difference in mean general vocabulary size among the four years ($F(3,292) = 6.420$, $p = .000$). Post-hoc analyses using Tukey HSD, shown in Table 5, further highlight significant differences between the 2018-2019 and 2019-2020 academic years, as well as the 2018-2019 and 2022-2023 academic years.

Regarding academic vocabulary size, Levene's test indicated unequal variances ($F(3,290) = 2.810$, $p = .040$). A Kruskal-Wallis test revealed a statistically significant difference in academic vocabulary size across the different years ($\chi2(3) = 30.734$, $p = .000$). Pairwise comparison, as detailed in Table 6, confirmed the significant differences observed in general vocabulary size.

To assess whether the observed differences could be attributed to the broader participant selection in the 2018-2019 academic year, which included substantially more experienced students, a comparison was made focusing solely on incoming students ($n = 192$). For general vocabulary size, Levene's test indicated equal variances ($F(3,188) = 0.170$, $p = .917$). Subsequently, a one-way ANOVA revealed a significant difference in

Table 5 Significance of general vocabulary size per academic year (p-values) (*=significant)

	2018-2019	2019-2020	2020-2021
2019-2020	0.003*		
2020-2021	0.998	0.075	
2022-2023	0.007*	0.911	0.176

Table 6 Significance of academic vocabulary size per academic year (p-values) (*=significant)

	2018-2019	2019-2020	2020-2021
2019-2020	0.017*		
2020-2021	0.644	1	
2022-2023	0.000*	1	0.231

Table 7 Significance of general vocabulary size per academic year for incoming students (p-values) (*=significant)

	2018-2019	2019-2020	2020-2021
2019-2020	0.050		
2020-2021	0.814	0.025*	
2022-2023	0.279	0.752	0.129

mean general vocabulary size across the four years ($F(3,188) = 3.850$, $p = .011$). Post-hoc analyses using Tukey HSD, see Table 7, showed only one significant difference between the 2019-2020 and 2020-2021 academic years.

For academic vocabulary size, Levene's test indicated unequal variances ($F(3,186) = 2.861$, $p = .038$). A Kruskal-Wallis test revealed no statistically significant differences in academic vocabulary size across the different years ($\chi2(3) = 7.769$, $p = .051$).

Overall, the initial significant differences between cohorts observed disappear when focusing solely on incoming students. This suggests that the observed differences could be attributable to the differing compositions of the participants groups on aspects like age or experience in higher education rather than to inherent disparities in vocabulary

level. As a result, the participants from all four academic years were combined in the subsequent analyses as a continuous student sample.

In order to determine whether the general and academic vocabulary screening tests perform similarly – showing comparable associations and effect sizes with personal and educational characteristics – a series of ANOVA tests and *t*-tests were conducted. As not all data were normally distributed, Spearman as well as Pearson correlation test were employed. Paired samples *t*-tests examined the association between the scores on the initial screening test to the retest.

To assess the extent to which an academic vocabulary size screening test predicts academic achievement, multiple regression models were calculated. Multivariate regression models were built to assess the impact of the two measures of vocabulary size and personal and educational factors. These were built gradually to detect whether decreases of the residual sum of squares (RSS) were significant ($p < 0.05$). A calculation of a variance inflated factor (VIF) score was performed to see if the two tests were collinear.

All data were processed using IBM SPSS version 24. For effect size, Cohen's guidelines were followed (1988).

5 Results

To assess the extent to which the performance of the academic vocabulary size screening test compares to that of the general vocabulary size screening test, the results of both tests were analyzed and put in relation to participants' personal and educational characteristics.

The average score on the academic vocabulary screening test ($n = 294$) was 82.74 on 100 ($SD = 7.72$) with a range of 43.08 (Min = 55.62, Max = 98.70). The average score on the general vocabulary screening test ($n = 296$) was 77.86 on 100 ($M = 77.86$, $SD = 5.75$) with a range of 36.75 (Min = 62.11, Max = 92.86).

A Spearman coefficient revealed a significant, weak, positive correlation between age and the scores on both vocabulary size screening tests (academic n = 288, $r = .179$, $p = .002$; general $n = 290$, $r = .168$, $p = .004$). Independent samples *t*-tests showed no significant difference in the score between men ($n = 77$, $M = 81.81$, $SD = 8.27$) and women ($n = 207$, $M = 83.00$, $SD = 7.61$) on the academic vocabulary screening test $t(282) = 1.136$, $p = .257$, or between men ($N = 77$, $M = 78.27$, $SD = 6.02$) and women ($N = 208$, $M = 77.81$, $SD = 5.68$) on the general one $t(283) = -.595$, $p = .552$. The effect size, as measured by Eta squared, showed no effect ($\eta^2 = 0.005$ and $\eta^2 = 0.001$ respectively). Regarding disability, independent samples *t*-tests revealed significant differences for both screening tests. On the academic vocabulary screening test, participants without a disability ($n = 209$, $M = 83.83$, $SD = 7.65$) outperformed those with a disability ($n = 65$, $M = 79.41$, $SD = 7.16$); $t(272) = -4.125$, $p = .000$ with a medium effect size ($\eta^2 = 0.06$). Similarly for general vocabulary,

participants without a disability (n = 209, M = 78,42, SD = 5.85) outperformed those with a disability (n = 66, M = 76.47, SD = 5.85); $t(273) = -2.420$, p = .016. The effect size was small (η^2 = 0.021).

An independent samples t-test found no significant difference in academic vocabulary size $t(273)$ = .611, p = .542 between participants who (also) spoke another language than Dutch at home (n = 38, M = 83.55, SD = 9.35) compared to participants who only identify Dutch as native language (n = 237, M = 82.72, SD = 7.47). Eta squared showed no effect (η^2 = 0.001). For general vocabulary size, there was a significant difference between participants who only or also identify another language than Dutch as their native language (n = 39, M = 76.01, SD = 6.57) compared to participants who only identify Dutch (n = 237, M = 78.30, SD = 5.58); $t(274) = -2.308$, p = .022, with a small effect size (η^2 = 0.019). Further analysis into native language through a one-way Anova found a significant effect $F(2,272) = 5.571$, p = .004 on academic vocabulary size with a small effect size (η^2 = 0.04). Tukey HSD post hoc test results revealed significant differences between Dutch as only native language (n = 237, M = 82.72, SD = 7.47) and also Dutch as native language (n = 24, M = 86.64, SD = 7.65) at p = .044, and between participants with also Dutch as a native language and those with only another different native language (n = 14, M = 78.24, SD = 9.81) at p = .003. For general vocabulary size, a one-way Anova revealed a significant difference among the three groups $F(2,273) = 9.382$, p = .000. The effect size (η^2 = 0.064) indicated a medium effect. Tukey HSD post hoc test results revealed significant differences between only another native language (n = 14, M = 71.66, SD = 5.77) and only Dutch (n = 237, M = 78.30, SD = 5.58) at p = .000 and with also Dutch as a native language (n = 25, M = 78.45, SD = 5.74) at p = .001.

An independent samples t-test showed a significant difference between students with no experience in higher education (n = 190, M = 80.71, SD = 7.39) and students with at least one semester of experience (n = 104, M = 86.46, SD = 6.91); $t(292) = -6.528$, p = .000. Eta squared showed a medium effect (η^2=0.127). An independent samples t-test also found a significant difference for participants without a degree (n = 276, M = 82.46, SD = 7.68) compared to participants with a degree (n = 16, M = 87.39, SD = 7.38); $t(290) = -2.502$, p =.013. Eta squared showed a small effect (η^2 = 0.021). For general vocabulary size, an independent samples t-test showed a significant difference between participants with no higher education experience (n = 192, M = 76.74, SD = 5.69) and participants with at least one semester experience (n = 104, M = 79.94, SD = 5.30); $t(294) = -4.745$, p = .000. The effect size indicated a medium effect (η^2=0.071). A significant difference was also found comparing participants without (n = 278, M = 77.64, SD = 5.73) and with a degree (n = 16, M = 82.11, SD = 4.82); $t(292) = -3.061$, p = .002. Eta squared showed a medium effect (η^2 = 0.071). A one-way Anova examining parental education found no differences among groups where only the mother (n = 42), only the father (n = 36), both (n = 114) or none (n = 78) had a higher education degree for both academic ($F(3, 265) = 1.522$, p = .209) and general vocabulary size ($F(3, 266) = .779$, p = .507). Both effects were trivial, η^2 = 0.017 and η^2 = 0.009 respectively.

Table 8 Effect sizes and significance for the personal and educational characteristics in relation to academic and general vocabulary size (*=significant,°=not significant)

	Academic			General			Differ?
Age	*	$r = .179$	S	*	$r = .168$	S	=
Gender	°	$\eta^2 = 0.005$	/	°	$\eta^2 = 0.001$	/	=
Native/multilingual	°	$\eta^2 = 0.001$	/	*	$\eta^2 = 0.019$	S	<
Dutch/multi/different	*	$\eta^2 = 0.04$	S	*	$\eta^2 = 0.06$	M	<
Disability	*	$\eta^2 = 0.06$	M	*	$\eta^2 = 0.021$	M	=
Experience in higher education	*	$\eta^2 = 0.127$	M	*	$\eta^2 = 0.071$	M	=
Already degree	*	$\eta^2 = 0.021$	S	*	$\eta^2 = 0.031$	S	=
Parental educational	°	$\eta^2 = 0.017$	S	°	$\eta^2 = 0.009$	/	>
Branches secondary education	*	$\eta^2 = 0.224$	L	*	$\eta^2 = 0.104$	M	>

Note: To indicate effect sizes, we used / for no effect ($\eta 2 \leq 0.01$ or $r \leq (-)0.1$), S for small effect ($0.01 < \eta 2 \leq 0.05$ or $(-)0.1 < r \leq (-)0.3$), M for medium effect ($0.05 < \eta 2 < 0.14$ or $(-)0.3 < r < (-)0.5$) and L for large effect ($\eta 2 \geq 0.14$ or $r \geq (-)0.5$) following Cohen's scales (1988).

A one-way Anova revealed a significant difference in academic vocabulary size among participants from different branches of Flemish secondary education ($F(3,282) = 27.165$, $p = .000$). The effect size was $\eta^2 = 0.224$, indicating a large effect. Tukey HSD post hoc test results revealed that there was a significant difference between Vocational Secondary School (BSO) ($n = 24$, $M = 75.89$, $SD = 6.44$) and all other levels, and between General Secondary School (ASO) ($n = 137$, $M = 86.40$, $SD = 6.29$) and Technical Secondary School (TSO) ($n = 116$, $M = 80.04$, $SD = 7.38$). Similar results were found for general vocabulary size. There was a significant difference among the branches $F(3,284) = 10.957$, $p = .000$ with a medium effect ($\eta^2 = 0.104$). Tukey HSD post hoc test results revealed significant differences between Vocational Secondary School (BSO) ($n = 24$, $M = 73.40$, $SD = 4.87$) and all other levels, and between General Secondary School (ASO) ($n = 138$, $M = 79.42$, $SD = 5.23$) and Technical Secondary School (TSO) ($n = 117$, $M = 76.93$, $SD = 5.76$). See Table 8 for an overview of effect sizes and significance for the personal and educational characteristics in relation to academic and general vocabulary size.

There was a strong correlation in academic vocabulary size between the scores of the initial screening test and the retest ($n = 71$, $r = .587$, $p = .000$). A paired samples t-test indicated that the score of the initial screening test ($M = 80.2085$, $SD = 6.90407$) did not significantly differ from the results of the retest ($M = 80.3075$, $SD = 8.19232$); $t(70) = -.120$, $p = .905$. The effect size, as measured by Cohen's d, was $d = 0.01$, indicating a trivial effect. The scores of the initial screening test of general vocabulary and the retest correlated moderately ($n = 71$, $r = .596$, $p = .000$). Results of a paired sam-

Table 9 Variance of GPA and CSE for general and academic vocabulary size (*=significant)

	General	Academic	Difference	Combined
GPA	R^2 = .032 adj R^2 = .026 $F(1,171)$ = 5.64, p = .019*	R^2 = .078 adj R^2 = .072 $F(1,170)$ = 14.33, p = .000*	R^2 = .046 adj R^2 = .046	R^2 = .084 adj R^2 = .073 $F(2, 169)$ = 7.839, p = .001*
CSE	R^2 = .031 adj R^2 = .025 $F(1,171)$ = 5.464, p = .021*	R^2 = .063 adj R^2 = .057 $F(1,170)$=11.395, p=.001*	R^2 = .032 adj R^2 = .032	R^2 = .071 adj R^2 = .060 $F(2, 169)$ = 6.421, p = .002*

ples *t*-test indicated that the score of general vocabulary of the initial screening test (M =77.2369, SD = 6.07683) significantly differed from the results of the retest (M = 79.0104, SD = 5.63255); $t(70) = -2.830$, p = .006. The effect size was d = 0.30, indicating a small effect.

Simple linear regression analysis was used to examine whether general and academic vocabulary size could account for variations in GPA and CSE of incoming students. As we only included incoming students starting higher education for the first time who participated in the initial January exam period, the final number of students included was n = 173 students. The regression results, as summarized in Table 9, revealed that vocabulary size explained between 2.5 % and 7.2 % of academic achievements. The impact was more pronounced for GPA. Comparing the explained variance of GPA and CSE between the general and the academic vocabulary size screening test, we observed a difference of 4.6 % and 3.2 %, respectively.

Multiple linear regression was used to examine how a model incorporating general and academic vocabulary size explains GPA and CSE. The fitted regression model was GPA=.728 + 0.039*general vocabulary size + 0.088*academic vocabulary size. The overall regression yielded statistical significance (R^2 = .084, $F(2, 169)$ = 7.839, p = .001). General vocabulary size did not significantly predict GPA (β = .039, p = .282) whereas the academic vocabulary size did (β = .088, p =.002). For CSE, the fitted regression model was CSE = −0.388 + 0.005*general vocabulary size + 0.009*academic vocabulary size. The overall regression was statistically significant (R^2 = .071, $F(2, 169)$ = 6.421, p = .002). General vocabulary size did not significantly predict CSE (β = .005, p = .235), academic vocabulary size did (β = .009, p = .008). To assess for multicollinearity, collinearity diagnostics were performed revealing a variance inflation factor [VIF] value below 5, indicating no presence of collinearity.

Further multiple linear regression additionally considered personal and educational characteristics. The model demonstrated no multicollinearity and no distinct indications

Table 10 Explained variance of GPA per model

Model	Coefficients B	SE	ß	*t*	*p*	R^2	ΔR^2	ΔR^2 sig.
Model 1						.021	.014	
Constant	5.618	2.846		1.974	.050			
General vocabulary size	.066	.037	.144	1.797	.074			
Model 2						.090	.078	.064
Constant	1.179	3.048		.387	.699			
General vocabulary size	.017	.039	.036	.429	.669			
Academic vocabulary size	.103	.030	.284	3.389	<.001			
Model 3						.133	.097	.019
Constant	3.724	3.201		1.163	.247			
General vocabulary size	.024	.040	.052	.591	.556			
Academic vocabulary size	.098	.031	.270	3.175	.002			
Age	–.045	.043	–.084	–1.064	.289			
Gender	–.792	.461	–.133	–1.718	.088			
Native language	–.708	.417	–.134	–1.698	.092			
Disability	.058	.469	.010	.123	.903			
Model 4						.217	.174	.077
Constant	11.399	3.632		3.139	.002			
General vocabulary size	.013	.039	.027	.325	.745			
Academic vocabulary size	.036	.033	.098	1.066	.288			
Age	–.015	.043	–.028	–.352	.725			
Gender	–.781	.444	–.132	–1.760	.080			
Native language	–.831	.410	–.157	–2.027	.044			
Disability	–.156	.452	–.026	–.346	.730			
Parental education	–.082	.348	–.019	–.237	.813			
Branch of secondary education	–.770	.195	–.350	–3.948	<.001			

of violations of homoscedasticity. The regression analysis included only students with complete data, resulting in a final sample of $n = 154$.

As shown in Tables 10 and 11, academic vocabulary added 6.4% of predictive value to GPA, personal factors 1.9%, and educational characteristics 8.1%. Academic vocabulary added 4.7% of predictive value to CSE, personal characteristics 2.6%, and educational characteristics 6.9%. Each change in R^2 was significant, as shown in Table 12.

Academic vocabulary size is the only significant variable in model 2 ($p = <.001$ and .004) for GPA and CSE, and in model 3 ($p = .002$) for GPA. For model 3 for CSE, academic vocabulary size is significant ($p = .013$) as is gender ($p = .023$). This significance of aca-

Table 11 Explained variance of CSE per model

Model	Coefficients B	SE	ß	*t*	*p*	R^2	ΔR^2	ΔR^2 sig.
Model 1						.019	.012	
Constant	.129	.318		.406	.686			
General vocabulary size	.007	.004	.137	1.711	.089			
Model 2						.072	.059	.047
Constant	−.304	.344		−.885	.378			
General vocabulary size	.002	.004	.043	.505	.614			
Academic vocabulary size	.010	.003	.249	2.932	.004			
Model 3						.121	.085	.026
Constant	−.092	.360		−.255	.799			
General vocabulary size	.004	.005	.072	.820	.414			
Academic vocabulary size	.009	.003	.215	2.510	.013			
Age	−.004	.005	−.073	−.912	.363			
Gender	−.120	.052	−.180	−2.306	.023			
Native language	−.051	.047	−.086	−1.084	.280			
Disability								
Model 4	.049	.053	.073	.923	.357	.198	.153	.068
Constant	.719	.411		1.750	.082			
General vocabulary size	.003	.004	.049	.570	.570			
Academic vocabulary size	.002	.004	.053	.566	.572			
Age	−.001	.005	−.021	−.265	.791			
Gender	−.118	.050	−.178	−2.344	.020			
Native language	−.065	.046	−.110	−1.402	.163			
Disability	.026	.051	.039	.508	.612			
Parental education	−.004	.039	−.009	−.107	.915			
Branch of secondary education	−.082	.022	−.333	−3.710	<.001			

demic vocabulary size is lost in model 4 for GPA and CSE (p = .288 and p = .572). In model 4 for both, the strongest significant variable is prior secondary education (p == <.001 and p = <.001). For GPA, native language gains significance (p = .044). For CSE, gender remains significant (p = .020).

Table 12 Significance per step of model for GPA and CSE

	GPA		CSE	
Model (n = 154)	***F* value**	***p* value**	***F* value**	***p* value**
Model 1	$F(1,152)=3.230$	.074	$F(1,152)=2.926$	.089
Model 2	$F(2,151)=7.468$	<.001*	$F(2,151)=5.835$	.004*
Model 3	$F(6,147)=3.748$	.002*	$F(6,147)=3.381$	.004*
Model 4	$F(8,145)=5.139$	<.001*	$F(8,145)=4.476$	<.001*

6 Discussion

This study aimed to explore whether an easy-to-administer, low-stakes, post-entry screening test of receptive academic vocabulary size can effectively predict early academic success of incoming students in a higher education context where Dutch is both the language of instruction and the predominant L1.

First, the results showed that the academic vocabulary size screening test showed slightly higher internal consistency and, in this context, unfolded similar associations, some with more pronounced effective sizes, with personal and educational characteristics of the participants compared to the general vocabulary size screening.

Both screening tests yielded high mean scores. The mean general vocabulary score in our study (77.86%) closely approaches the score Brysbaert et al. (2014) found for 60-year-olds. The main driver to which we attribute this discrepancy is level of education. Our study worked with a specific sample of students in higher education, which differs in level of prior education from the average population. Previous studies already showed that vocabulary size increases with higher levels of education (Brysbaert et al., 2014; Vandervieren & Casteleyn, 2020). Looking at this main driver, the academic vocabulary screening test also outperformed the general vocabulary screening test on prior education with respectively a large and medium effect. Our high mean scores seem not to be influenced by gender nor parental education, as this study did not find significant differences, diverging from results reported by other studies (Heeren et al., 2021; Warnby, 2023).

Interestingly, the mean score for the academic vocabulary test was even higher than that of the general vocabulary test. This finding contrasts with Masrai and colleagues (Masrai et al., 2021). We assume these elevated mean scores in our study overestimate students' actual academic vocabulary knowledge, particularly when the score is interpreted as 'total percentage of academic words known'. A possible explanation is that some items in the academic vocabulary test may not be sufficiently 'academic' and may

overlap with general vocabulary, especially considering that our participants are students who accessed higher education and predominantly speak the language of instruction as L1. Despite this high mean score, with a wider range of scores and a larger standard deviation compared to the general test, the academic vocabulary size screening test may more effectively differentiate proficiency within an L1 student cohort. Further research could look at finetuning the test by eliminating or replacing items to make sure they are infrequent enough (Nizonkiza & van Dyk, 2015).

Although we did not anticipate any significant vocabulary gains at the retest stage, a small but statistically significant improvement was found in general vocabulary size. No such gain was observed in academic vocabulary size. General vocabulary is more likely to be acquired incidentally in a linguistically rich environment like higher education, with regular exposure to diverse vocabulary through lectures, readings, and interactions. Academic vocabulary is lower in frequency, making it less accessible through incidental exposure alone (Vidal, 2011). This highlights the importance of deliberate instruction and targeted engagement to foster academic vocabulary growth.

The relationship between general and academic vocabulary size became more complex when examined through the lens of language background. Only the general vocabulary screening test showed a significant difference between monolingual Dutch speakers and multilingual students. However, when native language was classified into three categories – Dutch only, Dutch and another language, and only another language – both vocabulary tests yielded significant results, with the general vocabulary test showing a slightly stronger effect. Interestingly, participants who identified Dutch and another language as native languages outperformed those with only Dutch or only another language on the academic vocabulary test. This pattern did not emerge in the general vocabulary scores, suggesting a potential divergence in how multilingual students engage with academic versus general vocabulary. Nonetheless, these findings should be interpreted cautiously due to the relatively small subgroup sizes (25 multilingual, 14 non-Dutch mother tongue). These results highlight the nuanced challenges and differences faced by students with diverse linguistic backgrounds. The three-tiered classification used in this study (Dutch only, Dutch and another language, only another language) provides a more granular perspective compared to binary categorizations like native vs. non-native (Hazenberg & Hulstijn, 1996; Heeren et al., 2021) or Dutch vs. non-Dutch (Heeren et al., 2021; Welie at al., 2021). Further research could delve into the language profile and preparedness for higher education among students with diverse language backgrounds.

Second, the results showed relevant contributions of academic vocabulary size to account for variations in the two outcomes of academic achievement used, Grade Point Average (GPA) and Cumulative Study Efficiency (CSE). While disentangling the effects of a general and an academic vocabulary size screening test might be difficult (Masrai & Milton, 2018), academic vocabulary size more than doubles predictive power and even takes significance in the multiple linear regression models, where only the results of the

academic vocabulary size screening test significantly predict GPA and CSE. Approximately 5 to 6 % of the variance in GPA and CSE can be additionally explained by the academic vocabulary size screening test results. These findings align with existing literature on the moderate contribution of vocabulary to explaining the variance in academic outcomes and echo the additional contribution of academic vocabulary size to GPA predictability of 2 % to 7 % found in L2 research (Heeren et al., 2021; Masrai & Milton, 2018; Masrai & Milton, 2021; Townsend et al., 2012; Welie et al., 2021). Yet, the effect of the general vocabulary screening test was significantly stronger in that context. The combination of participants' L1 and educational background may have contributed to ceiling effects in the general vocabulary size screening test. This further supports the argument that an academic vocabulary screening test may be more appropriate and informative than a general vocabulary test in the context of higher education in L1. The results of the academic vocabulary size screening test continue to significantly predict both GPA and CSE scores when incorporating personal characteristics to the model. However, the significance is lost when including educational factors. A similar pattern emerges in a study by Heeren et al. (2021). As they put it, a measure like vocabulary "in a sample of mainly native speakers mostly reflects prior educational attainments" (Heeren et al., 2021). Academic vocabulary serves as a clear illustration of extended language cognition (Hulstijn, 2015), as it is not typically included in the everyday vocabulary of all speakers and is largely context-dependent – often first encountered in educational environments geared toward higher learning. Individuals with more advanced education are more likely to have come across and learned these less frequently used words. This underscores the need for intentional teaching and focused practice to effectively develop academic vocabulary, also in higher education.

It goes without saying that many factors impact academic achievement. In our most comprehensive model 82.6 % of GPA and 84.7 % of CSE remained unexplained, underscoring the complexity of the academic achievement variable (Heeren et al., 2021) where numerous other factors, such as attitude, intelligence, and personality which were not considered in this study, come into play (Davies, 2007). Yet, when dealing with complex outcome variables like academic success shaped by a range of contributing factors, even small effects, can have meaningful consequences (Cho & Bridgeman, 2012), especially if we can act on them.

This study faces several limitations. Firstly, the generalizability of the findings is limited due to the non-selective admission policy of the Flemish context, resulting in greater variability in academic preparedness and language proficiency compared to more selective systems. Secondly, the sample was confined to a single institution and initiative with voluntary participation. Additionally, the initiative was discontinued for older students after one year, leading to an overrepresentation in academic year 2018-2019. The design also allows for a larger deployment, particularly if a digital tool could be used for data collection. However, with high mean scores expressed as percentages, the results could convey a mixed message if the score is not reconsidered, for example adapting

to the success rate of different groups. A third limitation pertains to the choice of the format of a written vocabulary size screening. Although this matches well with our goal of establishing a quick academic vocabulary size test, it neglects spoken vocabulary, which may also impact early academic achievement. Insights on vocabulary depth could have provided additional nuanced insights in explaining specific outcomes, like the impact of language background.

7 Conclusion

The findings of this study highlight the potential value of an easy-to-administer academic vocabulary size screening test in a higher education context where the language of instruction is also the students' predominant L1. The screening test may offer a useful indication which incoming students might face challenges with academic language and obstacles in achieving academic success. The academic vocabulary size screening test outperforms the general alternative, with slightly stronger internal consistency, with similar associations, some with more pronounced effective sizes, with personal and educational characteristics of the participants, and with more predictive power for Grade Point Average (GPA) as well as Cumulative Study Efficiency (CSE) outcomes. However, our multiple linear regression models reveal that prior secondary education, limitedly supplemented by gender and language background, remains the most significant predictor of academic achievement, surpassing the impact of academic vocabulary size.

Regrettably, from an academic achievement perspective, personal and educational characteristics – such as prior education, gender, and language background – remain unalterable for students and higher education institutions. Academic vocabulary size of the instruction language however can be enhanced. By conducting a fast, low-stakes academic vocabulary size screening test for all incoming students, staff and institutions can raise awareness about the new linguistic demands and the impact on academic achievement without stigma.

Author contributions

Pieterjan Bonne: conceptualization, methodology, formal analysis, investigation, resources, data curation, writing – original draft, writing – review and editing, visualization; Jordi Casteleyn: conceptualization, methodology, writing – review and editing, supervision.

Funding information

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The work was conducted on a voluntary basis, with all time and resources contributed by the authors.

Statement of interest

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Statement of technology use

The authors used the following AI-based generative technologies: Microsoft Copilot was used for language and grammar revision. The authors certify that the output of this service was carefully checked, and they take full responsibility for the contents of this manuscript.

Supporting information

None.

References

Alderson, J.C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. A&C Black.

Alsahafi, M. (2023). The Relationship Between Depth of Academic English Vocabulary Knowledge and Academic Success of Second Language University Students. *SAGE Open*, *13*(1). https://doi.org/10.1177/21582440231153342

Al-Masrai, A. (2009). *Measuring the English vocabulary size of Saudi university students: Validating a new 10,000 word vocabulary size test*. (Unpublished master's thesis). Swansea University.

Bonne, P., & Casteleyn, J. (2021). Taalbeleid en taalondersteuning: Op zoek naar een gedeelde basis en strategie voor implementatie [Language policy and language support: In search of a shared basis and strategy for implementation.]. *Tijdschrift voor Onderwijsrecht en -Beleid*, *4*, 279-293.

Brysbaert, M., Keuleers, E., Mandera, P., & Stevens, M. (2014). *Woordenkennis van Nederlanders en Vlamingen anno 2013: Resultaten van het Groot Nationaal Onderzoek Taal* [Vocabulary of Dutch and Flemish people in 2013: Results of the Major National Language Survey]. Academia Press.

Cho, Y., & Bridgeman, B. (2012). Relationship of TOEFL iBT® scores to academic performance: Some evidence from American universities. *Language Testing*, *29*(3), 421-442. http://dx.doi.org/10.1177/0265532211430368

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates, Publishers.

Coxhead, A. (2000). A new academic word list. *TESOL quarterly*, *34*(2), 213-238. http://dx.doi.org/10.2307/3587951

Coxhead, A. (2016). Reflecting on Coxhead (2000), "A New Academic Word List". *TESOL Quarterly*, *50*(1), 181-185. http://dx.doi.org/10.1002/tesq.287

Daller, M., Müller, A., & Wang-Taylor, Y. (2021). The C-test as predictor of the academic success of international students. *International Journal of Bilingual Education and Bilingualism*, *24*(10), 1502-1511. http://dx.doi.org/10.1080/13670050.2020.1747975

Davies, A. (2007). Assessing Academic English Language Proficiency: 40+ Years of U.K. Language Tests. In J. Fox, M. Wesche, D. Bayliss, L. Cheng, C.E. Turner and C. Doe (Eds.), *Language Testing Reconsidered*, 73-86. University of Ottawa Press. http://dx.doi.org/10.2307/j.ctt1ckpccf.10

Deygers, B., Van den Branden, K., & Peters, E. (2017). Checking assumed proficiency: Comparing L1 and L2 performance on a university entrance test. *Assessing Writing*, *32*, 43-56.

De Wachter, L., & Heeren, J. (2013). Een taaltest als signaal. De ontwikkeling en implementatie van een strategische taalvaardigheidstoets [A language test as a signal. The development and implementation of a strategic language proficiency test]. *Levende Talen Tijdschrift*, *14*(1), 19-27.

Dunn, L.M., & Dunn, D.M. (2007). *Peabody Picture Vocabulary Test-Fourth Edition* (*PPVT-4*). APA PsycTests. https://doi.org/10.1037/t15144-000

Elder, C. (2017). Language Assessment in Higher Education. In E. Shohamy and N. Hornberger (Eds.), *Vol. 7 of the Encyclopedia of Language and Education*, 271-286. Netherlands: Springer. http://dx.doi.org/10.1007/978-3-319-02261-1_35

Feng, L., He, L., & Yiganmu, A. (2023). Determinants of students' academic success in English as a medium of instruction (EMI) classes: A systematic review. *Heliyon*, *9*(10). http://dx.doi.org/10.1016/j.heliyon.2023.e20421

Fonteyne, L., Duyck, W., & De Fruyt, F. (2017). Program-specific prediction of academic achievement on the basis of cognitive and non-cognitive factors. *Learning and Individual Differences*, *56*, 34-48. http://dx.doi.org/10.1016/j.lindif.2017.05.003

Geldof, D. (2014). Superdiversiteit als nieuwe realiteit: Oog voor veranderende integratiedynamieken [Superdiversity as a new reality: An eye for changing integration dynamics]. *Migratiemaatschappij: 20 stemmen over samenleven in diversiteit*, 33-50.

Glorieux, I., I. Laurijssen, & Sobczyk, O. (2014). *De instroom in het hoger onderwijs in Vlaanderen: Een beschrijving van de huidige instroompopulatie en een analyse van de overgang van secundair onderwijs naar hoger onderwijs* [The inflow into higher education in Flanders: A description of the current inflow population and an analysis of the transition from secondary education to higher education]. Steunpunt SSL.

Gu, L., & So, Y. (2015). Voices from stakeholders: What makes an academic English test 'international'?. *Journal of English for Academic Purposes*, *18*, 9-24. http://dx.doi.org/10.1016/j.jeap.2014.10.002

Hazenberg, S., & Hulstijn, J. (1996). Defining a Minimal Receptive Second-Language Vocabulary for Non-native University Students: An Empirical Investigation. *Applied Linguistics*, *17*(2), 145-163. https://doi.org/10.1093/applin/17.2.145

Heeren, J. (2024). Academic and general language proficiency in post-entry language assessment: Linguistic predictors of domestic L1 student achievement in the first year of university education. *ITL-International Journal of Applied Linguistics*, *175*(2), 271-297. http://dx.doi.org/10.1075/itl.23004.hee

Heeren, J., Speelman, D., & Wachter, L. De. (2021). A practical academic reading and vocabulary screening test as a predictor of achievement in first-year university students: implications for test purpose and use. *International Journal of Bilingual Education and Bilingualism*, *24*(10), 1458-1473.

Hulstijn, J. (2015). *Language Proficiency in Native and Non-native Speakers*. John Benjamins B.V.

Hyland, K. (2006). The 'other' English: thoughts on EAP and academic writing. *The European English Messenger*, *15*(2), 34-38.

Hyland, K., & Tse, P. (2007). Is There an "Academic Vocabulary"? *TESOL Quarterly*, *41*(2), 235-253.

Irvin, S., & Blankenship, K.G. (2022). Vocabulary & Academic Success in University Undergraduate Students. *Teaching and Learning in Communication Sciences and Disorders*, *6*(2). https://doi.org/10.30707/tlcsd6.2.1660595992.600476

Keuleers, M., Stevens, M., Mandera, P., & Brysbaert, M. (2015). Word knowledge in the crowd: measuring vocabulary size and word prevalence in a massive online experiment. *Q.J.Exp.Psychol. 68*, 1665-1692. https://doi.org/10.1080/17470218.2015.1022560

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language testing*, *16*(1), 33-51. https://doi.org/10.1177/026553229901600l

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*, 15-30.

Lemmouh, Z. (2010). *The relationship among vocabulary knowledge, academic achievement and the lexical richness in writing in Swedish university students of English*. Department of English, Stockholm University.

Masrai, A., & Milton, J. (2018). Measuring the contribution of academic and general vocabulary knowledge to learners' academic achievement. *Journal of English for Academic Purposes*, *31*, 44-57. https://doi.org/10.1016/j.jeap.2017.12.006

Masrai, A., & Milton, J. (2021). Vocabulary knowledge and academic achievement revisited: General and academic vocabulary as determinant factors. *Southern African Linguistics and Applied Language Studies*, *39:3* (August), 282-294. https://doi.org/10.2989/16073614.2021.1942097

Masrai, A., Milton, J., El-Dakhs, D.A.S., & Elmenshawy, H. (2021). Measuring the contribution of specialist vocabulary knowledge to academic achievement: disentangling effects of multiple types of word knowledge. *Asian-Pacific Journal of Second and Foreign Language Education*, *6*(1). https://doi.org/10.1186/s40862-021-00114-5

Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis* (pp. 57-78). Euro SLA.

Milton, J., & Treffers-Daller, J. (2013). Vocabulary size revisited: the link between vocabulary size and academic achievement. *Applied Linguistics Review*, *4*(1), 151-172. https://doi.org/10.1515/applirev-2013-0007

Nation, I.S.P. (1983). Testing and teaching vocabulary. *Guidelines*, *5*(1), 12-25

Nation, I.S.P. (2013). *Learning vocabulary in another language*. Cambridge University Press. https://doi.org/10.1017/CBO9781139858656

Nizonkiza, D., & van Dyk, T. (2015). Academic Literacy of South African Higher Education Level Students: Does Vocabulary Size Matter? *Stellenbosch Papers in Linguistics*, *44*, 147-174. https://doi.org/10.5774/44-0-159

Paribakht, T.S., & Webb, S. (2016). The relationship between academic vocabulary coverage and

scores on a standardized English proficiency test. *Journal of English for Academic Purposes*, *21*, 121-132.

Pecorari, D., Shaw, P., & Malmström, H. (2019). Developing a new academic vocabulary test. *Journal of English for Academic Purposes*, *39*, 59-71. https://doi.org/10.1177/0265532221145643

Peters, E. (2017). *Woordenschat aanleren in een vreemde taal. Hoe doe je dat?* [Learning vocabulary in a foreign language. How do you do that?]. ACCO.

Qian, D.D., & Lin, L.H.F. (2023). The Relationship Between Vocabulary Knowledge and Language Proficiency. *The Routledge Handbook of Vocabulary Studies*, 66-80. https://doi.org/10.4324/9780429291586-5

Read, J. (2015). *Assessing English Proficiency for University Study*. Palgrave, Macmillan.

Scheepers, R.A. (2019). Probing the depths: Can both size and depth of productive vocabulary predict academic performance?. *Per Linguam*, *34*(2). https://doi.org/10.5785/34-2-812

Schmitt, N. (2014). Size and Depth of Vocabulary Knowledge Size and Depth of Vocabulary Knowledge: What the Research Shows. *Language Learning*, *64*(4), 913-951.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language testing*, *18*(1), 55-88.

Skjelde, K., & Coxhead, A. (2020). Mind the gap: Academic vocabulary knowledge as a predictor of English grades. *Acta Didactica Norden*, *14*(3), 1-20. https://doi.org/10.5617/adno.7975

Townsend, D., Collins, P., & Biancarosa, G. (2012). Evidence for the Importance of Academic Word Knowledge for the Academic Achievement of Diverse Middle School Students. *The Elementary School Journal, 112*(3), 491-518. https://doi.org/10.1086/663301

Trenkic, D., & Warmington, M. (2019). Language and literacy skills of home and international university students: How different are they, and does it matter? *Bilingualism*, *22*(2), 349-365. https://doi.org/10.1017/S136672891700075X

Vandervieren, E., & Casteleyn, J. (2020). De relatie tussen taalcompetentie, nummerieke geletterdheid en academisch studiesucces: Een verkennende studie [The relationship between language competence, numerical literacy, and academic success: An exploratory study]. *Pedagogische Studiën*, *2020*(97(2)), 76-95.

van Kalsbeek, A., & Kuiken, F. (2014). Academisch taalgebruik in het Nederlandse wetenschappelijk onderwijs: stand van zaken en blik op de toekomst [Academic language use in Dutch scientific education: current state of affairs and future prospects]. *Internationale Neerlandistiek*, *52*(3), 221-236.

Van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension?. *Applied linguistics*, *34*(4), 457-479.

Vidal, K. (2011). A Comparison of the Effects of Reading and Listening on Incidental Vocabulary Acquisition. *Language Learning*, *61*(1), 219-258. https://doi.org/10.1111/j.1467-9922.2010.00593.x

Warnby, M. (2023). Academic vocabulary knowledge among adolescents in university preparatory programmes. *Journal of English for Academic Purposes*, *61*. https://doi.org/10.1016/j.jeap.2022.101203

Welie, C., Walet, L., Bernhards, S., & de Koster, M. (2021). De Academische Woordenschattoets (AWT): Een diagnostische toets bij de start van het hoger onderwijs [The Academic Vocabulary

Test (AWT): A diagnostic test at the start of higher education.] *Levende Talen Tijdschrift*, *22*(3), 30-40.

Wingate, U. (2015). *Academic literacy and student diversity: The case for inclusive practice* (Vol. 42). Multilingual Matters.

Appendix: Comparative overview

Table A.1 Comparative overview of elements of the results of academic and general vocabulary size screening and the relationship with personal and educational characteristics (*=significant)

	Academic	**General**
Test		
Cronbach's alpha	.91 (99/100 variables)	.88 (99/100 variables)
Real words	.94 (77/77 variables)	.90 (76/77 variables)
False words	.77 (22/23 variables)	.70 (22/23 variables)
Descriptive statistics		
M	82.74	77.86
SD	07.72	5.75
Range	43.08	36.75
Min	55.62	62.11
Max	98.70	92.86
Personal characteristics		
Age	p = .002*	p = .004*
Gender	p = .257	p = .552
Disability	p = .000*	p = .016*
Native/multilingual	p = .542	p = .022*
Only Dutch/also Dutch/only other	p = .004*	p = .000*
Only Dutch/also Dutch	p = .044*	p = .991
Only Dutch/Only other	p = .084	p = .000*
Also Dutch/Only Dutch	p = .003*	p = .001*
Educational characteristiscs		
Experience in higher education	p = .000*	p = .000*
Already degree	p = .013*	p = .002*
Parental education (Father/mother/both/none)	p =.209	p = .635
Branches of Flemish secondary education	p = .000*	p = .000*
– ASO/KSO	p = .728	p = .904
– TSO/KSO	p = .349	p = .193
– ASO/TSO	p = .000*	p = .002*
– ASO/BSO	p = .000*	p = .000*

 DOI 10.51751/dujal19780

Table A.1 Comparative overview (*cont.*)

	Academic	**General**
Test		
– KSO/BSO	p = .015*	p = .004*
– TSO/BSO	p = .037*	p = .022*
Retest		
Correlation with first test	p = .000*	p = .000*
Difference	p = .905	p = .006*

 DOI 10.51751/dujal19780