

## Teachers' receptive and productive vocabulary sizes in English-medium instruction

Hans Malmström, Diane Pecorari & Marcus Warnby

**To cite this article:** Hans Malmström, Diane Pecorari & Marcus Warnby (27 Sep 2023): Teachers' receptive and productive vocabulary sizes in English-medium instruction, Journal of Multilingual and Multicultural Development, DOI: [10.1080/01434632.2023.2260781](https://doi.org/10.1080/01434632.2023.2260781)

**To link to this article:** <https://doi.org/10.1080/01434632.2023.2260781>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 27 Sep 2023.



Submit your article to this journal [↗](#)



Article views: 951



View related articles [↗](#)



View Crossmark data [↗](#)

# Teachers' receptive and productive vocabulary sizes in English-medium instruction

Hans Malmström<sup>a</sup>, Diane Pecorari<sup>b</sup> and Marcus Warnby<sup>c</sup>

<sup>a</sup>Department of Communication and Learning in Science, Chalmers University of Technology, Gothenburg, Sweden; <sup>b</sup>School of Education, University of Leeds, Leeds, United Kingdom; <sup>c</sup>Department of Education and Special Education, University of Gothenburg, Gothenburg, Sweden

## ABSTRACT

Despite the growing popularity of English-medium instruction (EMI), the conditions for and consequences of teaching and learning academic content through English are poorly understood. The ability of teachers in the EMI environment (i.e. disciplinary or 'content' teachers) to engage students in English is central in this regard since intelligible interaction between the teacher and the students is a precondition for learning when the medium of instruction is English. Across EMI contexts, concerns have been raised about teachers' level of English proficiency (their ability to speak, write, read and listen in English), but research measuring their English proficiency attainments is lacking. This paper focuses on a key dimension of teachers' English proficiency: vocabulary knowledge. Teachers ( $n = 130$ ) took tests of receptive and productive knowledge of general and academic English vocabulary. The testing revealed significant proficiency variation in the cohorts tested, with some teachers exhibiting very low levels (<3000 words) of receptive and productive vocabulary knowledge. Implications for teaching in EMI are discussed.

## ARTICLE HISTORY

Received 26 July 2023  
Accepted 14 September 2023



## KEYWORDS

English-medium instruction; EMI; English proficiency; vocabulary knowledge; testing; teachers

## Introduction

Asked recently whether Englishisation – the increasing use of English-medium instruction (EMI) in higher education – is problematic, philosopher Philippe Van Parijs responded that it 'certainly is if ... the quality of education suffers badly as a result of transmission and interaction being hampered by a poor command of English by teachers and/or by students' (2021, 356). Van Parijs' response possibly speaks for many stakeholders in higher education (e.g. university management, teachers, and students), but the fact is that our knowledge regarding the nature and extent of the English challenges involved in EMI is limited.

To establish a stronger empirical basis in this regard, the present study shines a light on teachers in EMI and their English skills; research measuring the English proficiency attainments of such teachers is lacking. The paper focuses on one dimension of teachers' English proficiency, vocabulary knowledge, and takes this central dimension of communication to be indexical of teachers' broader linguistic competence (cf. Milton 2013).

**CONTACT** Hans Malmström  mahans@chalmers.se  Department of Communication and Learning in Science, Chalmers University of Technology, Hörsalsvägen 2, SE-412 962, Gothenburg, Sweden

This article has been corrected with minor changes. These changes do not impact the academic content of the article.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

## Literature review

In EMI settings, subject content is delivered in English (despite the fact that it is not the L1 for many, perhaps most participants), and it is frequently anticipated that the English language skills of participants will develop incidentally (Pecorari and Malmström 2018). While successful learning in EMI depends on a number of factors, the teacher's ability to communicate effectively in English is central. First, in terms of English language learning outcomes, teachers provide exposure to English. Second, in terms of teaching academic content, teachers require sufficiently strong English skills to be able to communicate clearly with their students, inside and outside of the classroom, in both speech and writing. Molino et al. (2023, 111) note that teachers' language use is critical for

ensuring the comprehension of disciplinary contents. Not only do lecturers define terms, explain concepts, and give examples, but they also redress misconceptions, guide students through discourse, make sure that learners focus on what is important, and establish meaningful interpersonal relations with them to facilitate the co-construction of meanings ... effective language use is also a matter of teacher credibility ... which may have an impact on the way students perceive English-medium instruction (EMI) courses and ultimately progress in their learning.

To date, research concerned with teachers' English proficiency in EMI has mainly relied on self-assessment measures involving teachers and students (rare counterexamples are studies of the Test of Oral English Proficiency for Academic Staff, e.g. Kling and Stæhr 2012, and the Interuniversity Test of Academic English, e.g. van Splunder et al. 2022). While some teachers express confidence about their ability to teach through an L2 (e.g. Jensen and Thøgersen 2011; Werther et al. 2014), many appear to 'underplay their limitations in general linguistic proficiency, for example searching for general or academic vocabulary, basic grammatical errors, or pronunciation issues' (Kling 2015, 218). Macaro draws a similar conclusion: 'one is left wondering whether [teachers] have fully grasped the level of competence needed to teach effectively. One dimension of that competence is their knowledge and expertise in their subject; the other dimension is their linguistic competence' (2018, 92). The possibility that teachers are not always reliable judges of their ability to teach EMI may explain why existing research has shown varied degrees of confidence in teacher proficiency (Campagna and Pulcini 2014; Hultgren, Jensen, and Dimova 2015; Margić and Vodopija-Krstanović 2017).

Despite the variability, the overall trend is concerning. In their review of research into teacher competency in EMI settings, Macaro et al. noted that 'more studies reported lecturers as identifying that they had linguistic problems than those that did not' (2018, 54). A similar conclusion is drawn in a recent systematic review of teachers' preparedness (Dang, Bonar, and Yao 2023). Importantly, for many teachers, it is teaching in English specifically that is experienced as onerous. In one study from Spain, 88.9% of teachers were confident about reading research literature in English, but many fewer (51.9%) felt that they had the language skills to teach in English (Fortanet-Gómez 2012).

Previous research has indicated that vocabulary is perceived as particularly problematic when teaching in English (e.g. Kling 2015; Tange 2010). Doiz, Lazagabaster, and Vázquez (2019) reported an illuminating example of this. In focus groups, EMI lecturers in Spain flagged insufficient English vocabulary knowledge as a key obstacle to teaching. A lecturer in History observed:

I am aware that as I go through the door in my class in English, my vocabulary goes from 5,000–1,000 words. In the classes in humanities like ours, this is an important problem because suddenly you see that you are much more 'limited' than in Spanish.

(2019, 158)

Similar concerns with vocabulary from teachers have been reported elsewhere, but relatively little research has sought to measure teachers' vocabulary size, and we are aware of no such research involving teachers in EMI settings.

The vocabulary knowledge teachers need depends on whether the nature of their tasks necessitates *general*, *academic* or *discipline-specific* vocabulary (Coxhead 2016), and whether it calls on *receptive* or *productive* knowledge of vocabulary. Teachers' varied tasks include listening and speaking to students in lectures, seminars, labs, etc.; reading research literature to inform their teaching;

reading and providing feedback on students' written work; writing teaching materials; and more (Lasagabaster and Doiz 2021). The rich research literature on lexical coverage in different domains (cf. Webb 2021) provides additional confirmation that receptive and productive vocabulary knowledge is critical for successful academic communication. There are indications that teachers are more challenged by the vocabulary outside their disciplinary field: 'many [teachers] will possess a very sophisticated English terminology in relation to their field of academic expertise and yet miss the words and phrases that enable them to engage in casual exchanges' (Tange 2010, 42). However, the relationship between the different types of vocabulary (knowledge) needed and attained by this group is poorly understood.

The present study addresses this gap in the research literature to verify stakeholders' perceptions (or misconceptions) about teachers' English proficiency. Based on the assumption that vocabulary knowledge is central for any kind of communicative competence – 'vocabulary breadth predicts overall language performance well' (Milton 2013, 65; see also Stæhr 2008) – it can be safely assumed that teachers' general and academic vocabulary knowledge is, at some level, indexical of their English proficiency. From this starting point, we seek first to establish the vocabulary knowledge attainments of teachers in an EMI setting. Given the broad range of functions indicated above which EMI requires teachers to undertake in English, the present study thus investigates both receptive and productive knowledge of general as well as academic vocabulary.

In this regard, the potential impact of three background variables is explored. The cognate effect in language learning has been amply demonstrated (e.g. De Groot and Keijzer 2000), and it affects measurements of English vocabulary size (Elgort 2013). In addition, gender differences have been identified in performance on vocabulary tests (e.g. Coxhead, Nation, and Sim 2014). Finally, it is reasonable to wonder whether teachers with more EMI teaching experience may be more proficient users of English (cf. Dafouz 2018). This study therefore considers the relationship between vocabulary knowledge and (i) first language (L1); (ii) gender; and (iii) experience in teaching in EMI contexts.

A further aim of this study is to address the question of whether teachers' English proficiency – as indexed by their vocabulary knowledge – provides them with a solid foundation for working in English-medium settings. This question is necessarily approached in a more exploratory manner than the first because, as noted above, the question of how much English is needed to teach in EMI does not yet have a satisfactory, evidence-based answer. However, while this question must be answered more tentatively, it is of critical importance; as noted by Molino et al. (2023, 123), 'providing more robust evidence of the features of [teacher language use] – and how they may vary – will enable stakeholders (lecturers, administrators, teacher trainers) to make decisions about the type of support needed for quality assurance in EMI'. This study therefore poses two questions:

1. What are the vocabulary knowledge attainments of teachers in EMI with respect to
  - a. first language (L1);
  - b. gender; and
  - c. experience in teaching in EMI contexts?
2. To what extent do those attainments provide preconditions for success in teaching in English?

## Methods

Addressing the questions articulated above entailed testing the general and academic vocabulary knowledge of teachers.

### Teacher participants

Participants in this study ( $n = 130$ ; 57 women, 70 men, and 3 who did not report their gender) were recruited through purposive sampling from three universities in Sweden, specifically from each university's foundational course for teaching in higher education (completing such a course is a formal requirement for anyone in a teaching role in Swedish higher education).

The majority of participants were PhD students, although some were in-service teachers in other roles. At the participating universities, PhD students make up a significant and important proportion of the teaching workforce (many doctoral students carry a teaching load similar to that of some senior faculty members). All three universities make extensive use of EMI (virtually all instruction at the master's level, and much at the undergraduate level, is conducted in English), and brand themselves as 'international' universities, and the participants would be expected to teach in English, including giving lectures, leading seminars, and/or overseeing laboratory work by students. The participants had varied durations of prior teaching experience (for analytical purposes, participants were divided into one of two 'experience' categories: those with less than one year's experience teaching in an EMI setting and those with one year of experience or more).

Approximately one quarter of the participants had Swedish as L1, while the remainder were L1 users of 28 different languages. The largest L1 groups were Germanic languages (other than English) (Swedish, German, and Dutch  $n = 47$ ); Romance ( $n = 23$ ); Chinese ( $n = 15$ ); and English ( $n = 10$ ). Information about the time of residency in Sweden was not gathered for the non-Swedes. To be admitted to their programmes, all participants had demonstrated a threshold level of proficiency in English. In the case of the Swedish speakers, this was done by means of having completed a level of secondary-school English considered to confer proficiency of at least the B2 level on the Common European Framework of Reference (CEFR), and for international participants, the equivalent on one of the internationally recognised tests.

Participation in the test sessions was voluntary (resulting in self-selection) and all participants gave their informed consent. Regulations in Sweden did not call for ethical approval for a study of this nature.

## Instruments

Receptive and productive (general and academic) vocabulary knowledge were measured using several instruments combined into a single test booklet and administered in testing sessions which lasted approximately 60 min. The testing was conducted in conjunction with a meeting of the higher education pedagogy class from which participants were recruited.

The test booklets included a brief background survey eliciting information about the participants' gender, first language and teaching experience. To measure *receptive general vocabulary knowledge*, selected parts of the Vocabulary Levels Test (VLT), version 2, were used (Schmitt, Schmitt, and Clapham 2001<sup>1</sup>). The VLT is a widely used monolingual matching-format test (see Figure 1 for an example item) providing an estimate of test takers' knowledge of words at the level of meaning recognition. Test takers match a word with the right meaning. For general vocabulary, the VLT is divided into four parts, each of which tests knowledge of 30 words at a given frequency band: the 2 K level (i.e. the 1001–2000 most frequent words in general English usage); the 3 K level (i.e. the 2001–3000 most frequent words); the 5 K level (i.e. the 4001–5000 most frequent words); and the 10 K level (i.e. the 9001–10,000 most frequent words). Because of a need to keep the test within approximately one hour, it was not possible to administer the full form of the VLT. For that reason, and because it was assumed *a priori* that the test takers in this study would have a reasonably high level of general vocabulary knowledge, the 2 K level was excluded. An aggregated VLT score of the three bands was used as an indicator of overall general vocabulary size (maximum 90 points).

The Academic Vocabulary Test (AVT; Pecorari, Shaw, and Malmström 2019) was used to test receptive knowledge of academic words. The AVT uses the same matching (meaning recognition) format as the VLT (see Figure 1), and tests 57 items (= maximum score) sampled from Gardner and Davies' (2015) Academic Vocabulary List (AVL). The AVT exists in two versions; Form 1 was used for this study.

To measure productive knowledge of general vocabulary, the Productive Vocabulary Levels Test (PVLT; Laufer and Nation 1999) was used. Like the VLT, it contains a sample of words representing

- (a)
- |   |           |       |         |
|---|-----------|-------|---------|
| 1 | brilliant | _____ | thin    |
| 2 | distinct  | _____ | steady  |
| 3 | magic     | _____ | without |
| 4 | naked     |       | clothes |
| 5 | slender   |       |         |
| 6 | stable    |       |         |
- (b)
- |       |                                |               |
|-------|--------------------------------|---------------|
| _____ | a promise to do something      | a. colleague  |
| _____ | a thing that somebody has made | b. commitment |
| _____ | someone who works with you     | c. creation   |
|       |                                | d. experiment |
|       |                                | e. flow       |
|       |                                | f. percentage |
- (c) He has a successful car \_\_\_\_\_ as a lawyer.
- (d) By aggr \_\_\_\_\_ the numbers they got a larger data set.

**Figure 1.** Example items from (a) the VLT (b) the AVT; (c) the PVLV; (d) the PAVT.

different frequency bands (18 items per frequency level); again, only the 3, 5 and 10 K sections were used. Items ( $3 \times 18$  for a maximum score of 54) came mainly from a single version of the PVLV, with a small number of replacements from other versions to avoid overlaps with the VLT. The PVLV measures controlled productive vocabulary knowledge, asking test takers to produce a target word deleted from a sentence based on a prompt (the initial letters of the target). So, for example, in Figure 1(c), the test taker should add the letters 'eer' to produce the target 'career'.

The Productive Academic Vocabulary Test (PAVT; Pecorari and Malmström 2019) was used to measure productive academic vocabulary knowledge. The PAVT uses the format of the PVLV to test knowledge of 52 items (= maximum score) from the AVL (as opposed to the University Word List used for the academic section of the PVLV). In the example from the PAVT shown in Figure 1(d), the test taker should add the letters 'egating' to produce the target 'aggregating'. Although validation of this test is ongoing, the PAVT has been extensively piloted in EMI environments and analysis in the present study confirms its reliability. The PAVT target words were not words which were tested on the AVT; however, they were selected on the same principles used in the construction of the AVT, including sampling words from a range of frequencies and a distribution of word classes which is representative of the composition of the AVL.

These instruments (i.e. receptive, general and academic; productive, general and academic) were combined into a single test paper with the background questionnaire coming between the receptive and productive part of the test. Table 1 provides an overview of the instruments used. Participants were told to work at a comfortable pace with no time limits, and that they should skip any question requiring them to guess blindly.

### Scoring

Binary scoring was applied to all tests. For the productive tests, a strict binary scoring principle was adopted, meaning that the answer had have been given entirely correctly with regard to spelling (standard regional variations accepted) and grammatical form.

### Analyses

To answer the first research question and its sub-questions, relating to the English vocabulary attainments of teachers and the three sub-samples (i.e. the participants grouped by gender, L1, and experience of teaching in English), statistical analyses were conducted in SPSS. Because several of the score distributions were non-parametric, both means and medians are reported as measures of central tendency, and non-parametric significance tests (Mann-Whitney and Kruskal-Wallis) were used when

**Table 1.** Overview of test instruments.

Vocabulary tested	Test instrument used	Format	No of words tested
Receptive, general	VLT version 2 (3, 5, and 10K)	Matching	30 per band x 3 = 90
Receptive, academic	AVT, form 1	Matching	57
Controlled productive, general	PVLT version A (3, 5, and 10K)	Produce prompted target	18 per band x 3 = 54
Controlled productive, academic	PAVT	Produce prompted target	52
		<b>Total words tested</b>	<b>253</b>

Note: Each section was analysed with respect to internal consistency using Cronbach's alpha (VLT  $\alpha = .95$ ; AVT  $\alpha = .90$ ; PVLT  $\alpha = .96$ ; PAVT  $\alpha = .95$ ).

comparing the between-group variation of the vocabulary scores among the three sub-samples. For general vocabulary only the aggregated scores for the VLT and PVLT were analysed.

Answering the second question, about preparedness for teaching in the English-medium environment, was less straightforward, for the reason that little is known about the degree of proficiency in English (or indeed another language) needed for teaching purposes. A CEFR level ranging from B2 to C2 is sometimes cited (e.g. Lasagabaster 2022), but such recommendations are apparently based on reasonable interpretations of the CEFR's can-do descriptors, rather than on empirical demonstration that they define necessary or sufficient knowledge. Lacking established benchmarks to answer this question, we adopted an exploratory approach informed by two concepts: *mastery thresholds*, i.e. the scores for a given frequency band which suggest that learners know all or most of the words at that level; and *lexical coverage*, i.e. estimates of what proportion of words must be known to perform particular communicative tasks.

The question of which scores indicate command of a given frequency band has been the subject of considerable discussion (e.g. Laufer 2021; McLean 2021). For example, for the VLT, Schmitt, Schmitt, and Clapham (2001) adopted 26/30 (or 87%). By contrast, Webb, Sasao, and Balance (2017) advocate frequency-sensitive thresholds, and suggest 29/30 (97%) up to the 3 K level, and 24/30 (80%) thereafter. Less discussion has attended the PVLT, but Laufer and Nation (1999, 41) acknowledge a degree of uncertainty, saying that 'satisfactory mastery of a level is a matter of judgement and depends what level is being considered, but is probably around 15 or 16 out of 18 ... for the 2000-word level'.

It is more tenuous to speak of mastery of academic vocabulary, which ranges across different frequencies. Schmitt, Schmitt, and Clapham (2001) discuss this with respect to their versions of the VLT, and the point is equally true for the AVT. However, a linkage study (Warnby, Malmström, and Yang Hansen 2023) has provided equivalences between the academic section of the VLT and the AVT, showing that a score of 26/30 or 29/30 on the academic section of the VLT corresponds to approximately 32/57 or 45/57 on the AVT.

In acknowledgement of the absence of general agreement on where to set mastery levels (or indeed how; cf. McLean 2021), we adopted a cautious approach, setting two benchmark points, one more generous and one more stringent. For receptive general vocabulary, those suggested by Schmitt, Schmitt, and Clapham (2001) and Webb, Sasao, and Balance (2017) were used, and then extended to receptive academic vocabulary, using the AVT scores which Warnby, Malmström, and Yang Hansen (2023) equate them with. For productive vocabulary (general and academic), the generous and stringent benchmarks were set with reference to the figures suggested by Laufer and Nation (1999) but lowered somewhat from the 5 K level, in line with Webb, Sasao, and Balance's (2017) considerations regarding frequency. Table 2 summarises the scores used as benchmarks in the analysis.

The discussion of mastery thresholds has concentrated on reading comprehension, and this is not the only skill needed by teachers. McLean argues for 'purpose-specific mastery thresholds' (2021, 129); yet, as noted above, too little is known about the English proficiencies needed by teachers in the EMI setting. For that reason, we do not assert that these scores indicate *mastery*, but regard them simply as benchmarks of *attainment*.

Coverage estimates refer to the approximate number of words needed in order to engage successfully in a communicative task. 'Success' is typically determined with reference to

**Table 2.** Attainment benchmarks.

Level	Receptive		Productive	
	Low (%)	High (%)	Low (%)	High (%)
3k	26/30 (87%)	29/30 (97%)	15/18 (83%)	16/18 (89%)
5K	24/30 (80%)	26/30 (87%)	14/18 (76%)	15/18 (83%)
10K	24/30 (80%)	26/30 (87%)	14/18 (76%)	15/18 (83%)
Academic	32/57 (56%)	45/57 (79%)	40/52 (76%)	43/52 (83%)

comprehension, and different coverage percentages for ‘acceptable’ and ‘optimal’ comprehension have been proposed in the literature, usually 95% and 98% respectively (cf. Laufer and Ravenhorst-Kalovski 2010). In exploring lexical coverage figures and vocabulary knowledge requirements, two kinds of communication which are relevant for teachers were considered in this study: (i) textbooks and (ii) seminars and lectures.

Hsu (2014) profiled engineering textbooks and found that knowledge of 5000 words provided 95% coverage, i.e. acceptable comprehension; for optimal comprehension (98% coverage), knowledge of 10000 words was needed. Based on transcript analyses of the British Academic Spoken English Corpus (160 lectures and 39 seminars from four disciplinary areas), Dang and Webb (2014) established that 95% coverage required knowledge of 4000 words whereas 8000 words were needed for 98% coverage. Dang (2022) profiled EMI lectures and found that 3000 and 7000 words were needed to reach 95% and 98% coverage in this teaching and learning context.

The vocabulary size estimates for acceptable and optimal comprehension cited above can be used as *indicative* benchmarks for determining if the teachers in this study can reasonably engage in relevant communicative tasks. It is important to note that the coverage figures are estimates of the vocabulary needed for *receptive* rather than *productive* purposes. However, as measures or estimates of the vocabulary students need for receptive purposes, they simultaneously index the vocabulary which educators *produce* in instructional settings. They therefore provide a picture of the vocabulary used by teachers when they produce these types of spoken academic discourse in English. While it is possible to communicate effectively with fewer lexical resources, these benchmark figures are indicative of the level of productive vocabulary knowledge which would allow teachers to perform in a similar way to their peers when interacting with students in seminars and lectures.

## Results

The first research question, regarding teachers’ English vocabulary attainments, we address by reporting the receptive and productive vocabulary test scores. Patterns related to L1 background, gender, and teaching experience can be seen, as well as considerable individual variation. We then turn to the second question, whether their level of vocabulary knowledge suggests preparedness to deliver university-level content in English.

### *What are teachers’ vocabulary knowledge attainments?*

Table 3 shows that the mean score for general vocabulary for all participants was 75.92/90 (84%) for receptive vocabulary and 31.55/54 (58%) for productive. For academic vocabulary, it was 47.38/57 (83%) for the receptive test, and for the productive test 25.96/52 (50%) (Appendix 1 presents the full results for all tests). The productive-receptive ratio for the full sample of 69% (for general vocabulary) and 60% (for academic vocabulary) falls within the range of 50% to 80% proposed by Milton (2009). The results reported in Table 3 reveal a large variation in vocabulary knowledge, as indicated by the standard deviations.

As indicated in Table 4, men scored consistently higher on each section of each test. At whole-test level, the differences were all significant. For general vocabulary, the median score on the



receptive test for men was 82 and for women 75 ( $U = 2580.50$ ,  $z = 2.84$ ,  $p = .005$ ,  $\eta^2 = .05$ ), and on the productive test it was 38 for men versus 27 for women ( $U = 2646.00$ ,  $z = 3.16$ ,  $p = .001$ ,  $\eta^2 = .08$ ). For academic vocabulary, median scores for men and women were 52 and 47 respectively ( $U = 2784.50$ ,  $z = 3.83$ ,  $p < .001$ ,  $\eta^2 = .12$ ), and 33 and 20 on the productive test ( $U = 2768.50$ ,  $z = 3.75$ ,  $p < .001$ ,  $\eta^2 = .11$ ). The gap between male and female teachers is larger for productive knowledge than receptive, and small at the 3 K level but then widening progressively.

Differences were also found when teachers' first language background was considered; see Table 5. Although participants had 28 different L1s, here we analyse only those from the four language families represented by 10 or more individuals: English ( $n = 10$ ); Germanic other than English ( $n = 47$ , including 33 L1 users of Swedish); Romance ( $n = 23$ ); and Chinese ( $n = 15$ ). On the receptive and productive tests, the L1 users of English and Germanic languages score higher than the overall mean ( $M_{VLT} = 84\%$ ;  $M_{AVT} = 83\%$ ;  $M_{PVLT} = 58\%$ ;  $M_{PAVT} = 50\%$ ), and in that order, with mean scores ranging from 95% to 67% for English and from 91% to 58% for the Germanic language speakers. The Romance L1 speakers score above the overall mean on the academic tests (AVT  $M = 88\%$ ; PAVT  $M = 58\%$ ), thus having a stronger score on the AVT than the Germanic L1 group. For general vocabulary, however, the Romance L1 group scores at the overall mean on the VLT ( $M = 84\%$ ) but below the mean on the PVLT ( $M = 53\%$ ). By contrast, the Chinese L1 speakers scored consistently below the mean on the general vocabulary tests, with 65% on the VLT and 29% on the PVLT, as well as on the academic tests, with 67% on the AVT and 23% on the PAVT. This same pattern holds true for all bands on the PVLT and nearly all on the VLT (at the 3 K level, the mean Germanic L1 scores were slightly higher than the English L1 scores). For each of the four tests (VLT, PVLT, AVT, PAVT), the four L1 groups were compared using a Kruskal–Wallis test. The results indicated a statistically significant difference for all tests,  $p < .001$ , with the following effect sizes,  $\eta^2_{VLT} = .39$ ,  $\eta^2_{AVT} = .24$ ,  $\eta^2_{PVLT} = .37$ , and,  $\eta^2_{PAVT} = .24$ . The post-hoc pairwise comparisons revealed significant differences between 14 of the 24 pairs, as indicated by Table 6. The largest differences were found between the L1 Chinese and all other groups (Appendix 2 provides a visualisation of the L1 group differences across the tests).

A further consideration was whether experience of teaching in EMI might be associated with English vocabulary knowledge. As Table 7 shows, save for the 3 K level, the mean scores are consistently higher in the more experienced group. However, the differences were not statistically significant with respect to general receptive and productive vocabulary knowledge and academic productive knowledge; the receptive academic vocabulary scores are statistically significantly higher for the participants with one or more years of EMI teaching experience ( $Mdn = 50.50$ ) than for those with less than one year's experience ( $Mdn = 48$ ), but the effect size is rather small,  $U = 2498.00$ ,  $z = 2.00$ ,  $p = .045$ ,  $\eta^2 = .03$ .

### Is the teachers' vocabulary knowledge fit for purpose?

A key question is whether the level of vocabulary knowledge demonstrated by these teachers is adequate for the task of teaching through the medium of English. As noted above, there is insufficient evidence to indicate what level of English proficiency EMI requires (much less which proficiencies).

**Table 3.** Full sample test results.

	Receptive vocabulary tests					Productive vocabulary tests				
	General				Acad AVT	General				Acad PAVT
	3K	5K	10K	Tot VLT		3K	5K	10K	Tot PVLT	
Test max. points	30	30	30	90	57	18	18	18	54	52
<i>M</i>	28.58	27.24	20.1	75.92	47.38	11.85	11.02	8.68	31.55	25.96
<i>M</i> %	95%	91%	67%	84%	83%	66%	61%	48%	58%	50%
<i>SD</i>	2.39	3.65	7.98	12.86	7.65	4.03	4.60	5.06	12.99	13.23
<i>Mdn</i>	30	29	21	79	49	13	11	9	33.50	26

In this section, we therefore consider teachers' vocabulary knowledge against reference points from the literature which may provide a useful perspective: attainment benchmarks and lexical coverage estimates.

Table 8 presents the 10<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup>, 75<sup>th</sup>, and 90<sup>th</sup> percentiles for each section and band on the test, and considerable variation in the scores can be seen. For example, the top decile group reached the high benchmark in nearly all tests, with their median score falling short - by .1 point - only on the PAVT. Similarly, the top quartile met one of the benchmarks on all sections except the PAVT and the productive 10 K. By contrast, the bottom quartile reached the low benchmark only for the receptive 3 and 5 K and AVT, and the bottom decile reached the low benchmark only for the AVT. In hindsight, it is regrettable that time constraints did not permit including the 2K-level in the VLT and PVLt; this would have provided a more granular assessment of the vocabulary knowledge of the weakest test takers.

Although there exists no empirically validated threshold for the proficiencies needed to teach in English, the findings presented here give reason to conclude tentatively that some teachers are likely to be underprepared because, as Stæhr remarked, those who 'do not know the most frequent 2000 or 3000 words in English ... will have severe difficulties in understanding most written and spoken text and it will make it even more difficult to engage actively in written and spoken communication' (2008, 150). Thus, it is reasonable to suspect that the 50% with limited productive knowledge of the 3000 most common words in English may struggle to lecture cogently on complex topics; and to suppose that the 10% who have limited receptive knowledge at the 3 K level may find it difficult to understand students' questions, or to understand fully, and assess fairly, their written work.

The teachers' scores on the receptive vocabulary tests can be benchmarked against the vocabulary loading of several relevant tasks. Hsu's (2014) profile of engineering textbooks indicates that knowledge of 5000 words is needed to achieve 95% coverage (adequate comprehension), while the 10 K level is required for 98% coverage (optimal comprehension). The bottom decile does not reach even the lower threshold. The bottom quartile falls between the lower and higher benchmark levels for the 5 K (indicating adequate but not optimal comprehension), while the remainder of the test-takers exceed both benchmark scores. In other words, the majority of teachers tested would experience few problems when engaging with textbooks, but a minority could be challenged. The impact on that minority is uncertain, but understanding the textbook content is arguably crucial for effective curriculum alignment, lesson planning, and supplementing instruction in lectures and seminars.

A second point of comparison is engagement with students during seminars; comprehension is critical since it allows teachers to assess students' understanding of the course content, facilitate learning, encourage participation, address individual needs, build relationships, and monitor progress. Dang and Webb's (2014) research indicates that 4,000 and 8,000 words are needed for 95% and 98% coverage. Again, the bottom decile falls below both attainment benchmarks at the 3 K level. Notably, only the top quartile of test takers has receptive vocabulary knowledge suggesting optimal (98%) comprehension.

Turning to productive vocabulary, and the teachers' ability to lecture, lead seminars, etc., there is greater cause for concern. Although such interactions involve a greater proportion of high-frequency vocabulary than written academic texts, the participants demonstrated much lower productive knowledge. According to Dang and Webb (2014) and Dang (2022) the 95% coverage level of lecture/seminar context can be achieved with knowledge of about 3000–4000 words (the lower number applying to EMI lectures). Only the top quartile appears to know that many words productively. To reach the higher 98% coverage figure, knowledge of between 4000 and 8000 words is needed. The VLT does not measure the intermediary levels between 5 K and 10 K. However, the bottom 75% who did not master the 3 K level can be confidently assumed not to master the higher levels; and only the top quartile, who did reach the 10 K attainment benchmark, can be confidently assumed to master the intermediary levels.

It should be reiterated that the coverage figures used here as reference points were developed to understand how much vocabulary students need to reach good comprehension of teaching

**Table 4.** Sub-sample test results (women vs. men).

		Receptive vocabulary tests					Productive vocabulary tests				
		General				Acad AVT	General				Acad PAVT
		3K	5K	10K	Tot VLT		3K	5K	10K	Tot PVLТ	
Test max. points		30	30	30	90	57	18	18	18	54	52
Women	M	28.28	26.47	18.46	73.21	44.89	10.96	9.60	7.04	27.60	21.16
	M %	94%	88%	62%	81%	79%	61%	53%	39%	51%	41%
	SD	2.68	3.82	7.36	12.43	7.25	4.08	4.28	4.67	12.34	11.90
	Mdn	29	28	19	75	47	12	10	6	27	20
Men	M	28.86	27.89	21.37	78.11	49.44	12.54	12.11	10.01	34.67	29.96
	M %	96%	93%	71%	87%	87%	70%	67%	56%	64%	58%
	SD	2.09	3.34	8.15	12.6	7.26	3.93	4.48	4.99	12.66	12.97
	Mdn	30	29	23	82	52	13	13	11	38	33

**Table 5.** Sub-sample test results (L1 groups).

		Receptive vocabulary tests					Productive vocabulary tests				
		General				Acad. AVT	General				Acad. PAVT
		3K	5K	10K	VLT		3K	5K	10K	PVLT	
Test max. points		30	30	30	90	57	18	18	18	54	52
Full sample M %		95%	91%	67%	84%	83%	66%	61%	48%	58%	50%
L1	<i>M</i>	29.4	29.6	26.8	85.8	51.9	15.1	14.7	12.4	42.2	35
English	M %	98%	99%	89%	95%	91%	84%	82%	69%	78%	67%
	<i>SD</i>	1.27	0.97	5.72	7.86	7.33	3.64	4.14	3.84	11.33	13.97
	<i>Mdn</i>	30	30	28.5	88.5	55	16	16	12.5	45.5	37.5
	<i>10th</i>	26.30	27.20	12.50	66.10	34.50	6.6	5.5	4.5	16.6	7
	<i>25<sup>th</sup></i>	29	29.75	26.75	85.75	48.75	14.25	13.75	10.5	38.5	28.75
L1 other	<i>M</i>	29.7	28.64	23.85	82.19	49.23	13.40	13.15	10.79	37.34	30.19
Germanic	M %	99%	95%	80%	91%	86%	74%	73%	60%	69%	58%
	<i>SD</i>	0.62	1.96	5.54	7.36	6.14	2.94	3.50	4.63	10.04	11.40
	<i>Mdn</i>	30	29	25	85	50	14	13	12	39	31
	<i>10th</i>	29	24.80	15	69.60	39	8	8	3.8	20.6	12.4
	<i>25<sup>th</sup></i>	30	28	21	78	46	12	10	7	28	24
L1	<i>M</i>	28.39	27.04	20.00	75.43	50.22	10.52	10.09	8.22	28.83	27.87
Romance	M %	95%	<b>90%</b>	67%	84%	88%	<b>58%</b>	<b>56%</b>	<b>46%</b>	<b>53%</b>	<b>54%</b>
	<i>SD</i>	1.75	2.27	4.88	7.65	5.21	3.69	3.53	3.92	10.36	10.38
	<i>Mdn</i>	29	27	20	76	51	11	10	7	26	29
	<i>10th</i>	25	23.4	13.20	64.60	42	5.8	6	3.4	17.2	17.2
	<i>25<sup>th</sup></i>	28	25	16	70	47	7	8	5	22	22
L1	<i>M</i>	25.93	22.2	10.2	58.33	38.27	7.53	5.33	3.00	15.87	12.13
Chinese	M %	<b>86%</b>	<b>74%</b>	<b>34%</b>	<b>65%</b>	<b>67%</b>	<b>42%</b>	<b>30%</b>	<b>17%</b>	<b>29%</b>	<b>23%</b>
	<i>SD</i>	3.79	5.36	7.19	13.85	8.06	3.64	3.81	3.95	10.89	12.15
	<i>Mdn</i>	27	23	10	58	38	6	4	2	14	9
	<i>10th</i>	19.40	15.20	1.60	38	25.40	3.8	1.6	1	6.2	1
	<i>25<sup>th</sup></i>	23	17	4	48	34	5	3	1	9	3

Note 1. Percentages in bold indicate a sub-sample mean below the full sample mean.

Note 2. 10th indicates bottom decile; 25th indicates bottom quartile.

**Table 6.** Post-hoc pairwise comparisons between language groupings.

Comparison pairs	VLT	PVLT	AVT	PAVT
Chinese-English	***	***	***	***
Chinese-Germanic	***	***	***	***
Chinese-Romance	.06	.10	***	*
English-Germanic	.81	1	.7	1
English-Romance	**	**	1	.33
Germanic-Romance	*	**	1	1

\*\*\*  $p > .001$ , \*\*  $p > .01$ , \*  $p > .05$ . Significance values have been adjusted by the Bonferroni correction for multiple tests.

**Table 7.** Sub-sample test results (EMI teaching experience).

	Receptive vocabulary tests					Productive vocabulary tests					
	General				Acad. AVT	General				Acad. PAVT	
	3K	5K	10K	Tot VLT		3K	5K	10K	Tot PVLT		
Test max. points	30	30	30	90	57	18	18	18	54	52	
<1 yr EMI teach exp	<i>M</i>	28.87	27.13	19.89	75.89	46.41	11.54	10.98	8.31	30.84	24.25
	<i>M %</i>	96%	90%	66%	84%	81%	64%	61%	46%	57%	47%
	<i>SD</i>	1.97	3.45	8.26	12.58	7.19	3.8	4.4	4.84	12.32	12.35
	<i>Mdn</i>	30	28	21	79	48	12	11	8	31	24
≥1 yr EMI teach exp	<i>M</i>	28.4	27.44	20.49	76.32	48.47	12.18	11.15	9.12	32.44	27.82
	<i>M %</i>	95%	91%	68%	85%	85%	68%	62%	51%	60%	54%
	<i>SD</i>	2.67	3.77	7.66	12.93	7.8	4.25	4.75	5.21	13.53	13.68
	<i>Mdn</i>	30	29	21	79.5	50.5	13	12.50	10.50	35	31

**Table 8.** Scores by percentile in relation to attainment benchmarks.

		Percentiles				
		10	25	50	75	90
Receptive	3K	25	28*	30**	30**	30**
	5K	23	25*	29**	30**	30**
	10K	8	15	21	28**	29**
	AVT	37.10*	42*	49**	54**	56**
Productive	3K	6	9	13	15*	17**
	5K	4	7.75	11	15*	17**
	10K	1.10	4	9	13	15**
	PAVT	8	14.75	26	36.25	42.90*

\*Reaches the low attainment benchmark.

\*\*Reaches the high attainment benchmark.

interactions, not to define how many words teachers need to conduct them. They are meaningful, though, in that they indicate that the teachers represented in the spoken corpora used in these studies have greater lexical resources to conduct lectures, tutorials, etc., than the teachers in the present study.

## Discussion

Teacher proficiency in the medium of instruction is important and can affect factors as diverse as teachers' confidence and their classroom behaviours (Dewaele and Leung 2022). The present study measured one aspect of proficiency, the general and academic vocabulary knowledge of teachers in or soon destined for EMI. Unsurprisingly, these teachers performed much more strongly on the receptive tests, and this confirms earlier research (e.g. Fortanet-Gómez 2012) that has pointed to fewer perceived problems among teachers when it comes to receptive as opposed to productive tasks.

That Germanic-language speakers had an advantage is no doubt attributable in part to the cognate effect. The relatively strong performance of the Romance speakers may be due to the significant influences of Romance languages on the English lexicon, and of Greco-Latin origins on the

academic word list.<sup>2</sup> Similarly, the below-average performance of the Chinese speakers is likely due in part to their L1 being typologically very different from English.

The finding that men consistently outperformed women is noteworthy, given contradictory findings emerging from previous research on students in EMI. Coxhead, Nation, and Sim (2014) found that women outperformed men in general vocabulary. By contrast, Green et al. (2023) reported no significant difference for general vocabulary, but did find that male students outperformed females in terms of receptive academic vocabulary knowledge.

The absence of a difference between more and less experienced teachers is somewhat surprising. It would be natural to assume that the additional engagement with English that follows from working longer in an English-medium context caused teachers to develop their English vocabulary knowledge (along with other dimensions of English proficiency), just like some teachers seem to be hoping for (e.g. Dafouz 2018). The difference in teaching experience (<1 year vs. ≥ 1 year) may be too small to detect additional possible effects (beyond academic vocabulary) which experience may have on vocabulary development (since vocabulary development is a slow process with a linear relationship between word frequency and development).

These factors contribute to explaining the wide variation in scores which was observed and was presented in Table 3. For instance, although a ceiling effect applied to or was approached by the top decile on virtually all sections of all tests, scores for the bottom decile are low, often worryingly so: for receptive, general vocabulary (all levels), the median score was 63%; at the 10 K only 26.6%; and on the AVT, 55.6%. For productive vocabulary, at the 3 K level, the bottom decile scored 33.3%, and for academic vocabulary only 23.3%.

'Only', in this context, is of course a somewhat problematic characterisation, because evidence for how much English vocabulary is needed for pedagogical tasks is lacking. However, to the extent that findings about the vocabulary needed for other tasks, such as reading a textbook or engaging in lecture discourse, can provide a relevant reference point, it is clear that while some individuals are very well placed for the task, others are likely to struggle, and if the teachers struggle, it is unlikely that their students will not be impacted.

## Conclusion

The pan-European survey of English-taught programmes undertaken by Wächter and Maiworm (2014) reported that a very large majority (95%) of the programme directors surveyed believed that the English proficiency of their teachers assigned to EMI duty was 'good' or 'very good'. However, it is abundantly clear that the participants in this study will be going into the EMI classroom with greatly diverse abilities to communicate in English. This indicates that it is not possible to generalise about the linguistic preparedness of teachers in EMI settings; some will be prepared, others much less so. The high degree of complexity surrounding the issue of teachers' English proficiency – and ways of assessing it – has been confirmed by research in other Nordic EMI teaching contexts (e.g. Dimova and Kling 2015) as well as in other places of EMI implementation (e.g. Dang, Bonar, and Yao 2023).

Both of these observations, i.e., that teachers' vocabulary knowledge is varied, and in some cases possibly insufficient, are likely to be true of the broader EMI context in Sweden (and perhaps other contexts); though, as noted, valid, reliable and objective assessment of teachers' English proficiency has been lacking (Macaro et al. 2018). In Sweden, English in higher education is regarded as a relatively neutral lingua franca, and little attention is given to stakeholders' proficiency in English. It is a routine if questionable assumption that a degree from a university in an Anglophone country, a masters' or PhD dissertation written in English, or authorship of research publications in English indicate preparation to teach in English. Typically, no systematic evaluation of prospective teachers' English proficiency is made, and because English is only a minor consideration in determining who becomes a teacher, varied and, in some cases, insufficient levels of proficiency are virtually inevitable. Regrettably, even when teachers' English proficiency is found wanting, systematic teacher

development programmes are rare and, when they are available, far from always succeed in addressing the specific, and in many cases varied, needs of teachers in EMI (Lasagabaster 2022).

Several implications of this situation deserve consideration, the first of which is the impact on student learning outcomes. If some teachers struggle to communicate effectively in English, it must impair their ability to deliver course content, write clear and unambiguous instructions for assessments, and in other ways interact with students, and that can only have a negative effect on student learning. Obviously, teaching is an immensely complex activity, and many factors other than English proficiency determine success or failure; thus, teachers with insufficient English proficiency can deploy compensatory strategies, but it is difficult to see how they could entirely neutralise negative effects on classroom interaction.

Diminished content learning is not the only potentially negative outcome. Assuming that teachers can marshal their lexical resources in order to present content in an effective way, the use of a relatively simplified vocabulary may call into question the teacher's credibility, which Molino et al. (2023) correctly point out as being of importance. In addition, the productive scores indicate that teachers use, and therefore provide exposure to, relatively little mid- and low-frequency vocabulary, limiting their role in promoting students' English vocabulary acquisition, contrary to expectations. In at least some EMI settings around the world, the development of students' English proficiency is an expected outcome, and teachers supply some of the critical exposure needed to foster that development (Rose et al. 2020). If the quality of that English exposure is inadequate, the prospects for English language development will also diminish.

These implications demonstrate vividly the need for awareness amongst policy makers, administrators, and others involved in decision-making related to the implementation of EMI to understand which teachers are prepared for EMI, which are not, and which are in need of support (cf. Dang, Bonar, and Yao 2023). To support such informed decision-making, though, there is also a demand for more evidence speaking to both sides of the question: which English-language proficiencies teachers need, and to what extent they have them; as well as the interrelatedness of pedagogical and linguistic preparation, i.e. to what extent can good pedagogical skills compensate for linguistic limitations, and vice versa?

The absence of evidence-based answers to those questions places limitations on this study. We have discussed whether the teachers in this study have the level of vocabulary knowledge needed for success in their pedagogical roles while acknowledging that neither we nor anyone else knows what that level is (Lasagabaster 2022). To this end, future research could help establish reasonable English proficiency attainment benchmarks for teaching in the English-medium environment, because as the expansion of EMI continues unabated continued attention to the teachers who deliver it is needed.

## Notes

1. The widely used versions of the VLT by Schmitt, Schmitt, and Clapham et al. were based on the work of Paul Nation (e.g., Nation 1983).
2. This interpretation is borne out by the strong performance observed by five Greek speakers on the academic vocabulary test although they were too small a group numerically to make statistical analysis meaningful.

## Acknowledgement

The financial support of the Torsten Söderberg Foundation (Ö25/21) is gratefully acknowledged. As is the logistical support of the course conveners enabling the data collection at the three collection sites.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## References

- Campagna, S., and V. Pulcini. 2014. "English as a Medium of Instruction in Italian Universities: Linguistic Policies, Pedagogical Implications." *Textus, English Studies in Italy* 1: 173–190. <https://doi.org/10.7370/77492>.
- Coxhead, A. 2016. "Acquiring Academic and Disciplinary Vocabulary." In *The Routledge Handbook of English for Academic Purposes*, edited by K. Hyland, and P. Shaw, 177–190. Abingdon: Routledge.
- Coxhead, A., P. Nation, and D. Sim. 2014. "Creating and Trialling Six Versions of the Vocabulary Size Test." *TESOLANZ Journal* 22: 13–27.
- Dafouz, E. 2018. "English-Medium Instruction and Teacher Education Programmes in Higher Education: Ideological Forces and Imagined Identities at Work." *International Journal of Bilingual Education and Bilingualism* 21 (5): 540–552. <https://doi.org/10.1080/13670050.2018.1487926>.
- Dang, T. N. Y. 2022. "Vocabulary in Academic Lectures." *Journal of English for Academic Purposes* 58: 1–10. <https://doi.org/10.1016/j.jeap.2022.101123>.
- Dang, T. K. A., G. Bonar, and J. Yao. 2023. "Professional Learning for Educators Teaching in English-Medium-Instruction in Higher Education: A Systematic Review." *Teaching in Higher Education* 28 (4): 840–858. <https://doi.org/10.1080/13562517.2020.1863350>.
- Dang, T. N. Y., and S. Webb. 2014. "The Lexical Profile of Academic Spoken English." *English for Specific Purposes* 33: 66–76. <https://doi.org/10.1016/j.esp.2013.08.001>.
- De Groot, A. M. B., and R. Keijzer. 2000. "What Is Hard to Learn Is Easy to Forget: The Roles of Word Concreteness, Cognate Status, and Word Frequency in Foreign-Language Vocabulary Learning and Forgetting." *Language Learning* 50: 1–56. <https://doi.org/10.1111/0023-8333.00110>.
- Dewaele, J.-M., and P. P. Y. Leung. 2022. "The Effect of Proficiency on non-Native English as a Foreign Language (EFL) Teachers' Self-Efficacy and Practice in the EFL Classroom." *IAFOR Journal of Education* 10: 11–32. <https://doi.org/10.22492/ije.10.1.01>.
- Dimova, S., and J. Kling. 2015. "Lecturers' English Proficiency and University Language Policies for Quality Assurance." In *Integrating Content and Language in Higher Education: From Theory to Practice Selected Papers from the 2013 ICLHE Conference*, edited by R. Wilkinson, and M. Walsh, 50–65. Bern: Peter Lang.
- Doiz, A., D. Lasagabaster, and V. P. Vázquez. 2019. "The Integration of Language and Content in English-Medium Instruction Courses." *Ibérica* 38: 151–176.
- Elgort, I. 2013. "Effects of L1 Definitions and Cognate Status of Test Items on the Vocabulary Size Test." *Language Testing* 30: 253–272. <https://doi.org/10.1177/0265532212459028>.
- Fortanet-Gómez, I. 2012. "Academics' Beliefs about Language Use and Proficiency in Spanish Multilingual Higher Education." *ALLA Review* 25: 48–63. <https://doi.org/10.1075/aila.25.04for>.
- Gardner, D., and M. Davies. 2015. "A New Academic Vocabulary List." *Applied Linguistics* 35: 305–327. <https://doi.org/10.1093/applin/amt015>.
- Green, C., M. Pantelich, M. Barrow, D. Weerasinghe, and R. Daniel. 2023. "Receptive Vocabulary Size Estimates for General and Academic Vocabulary at a Multi-Campus Australian University." *Australian Review of Applied Linguistics* 1–21, advance online publication. <https://doi.org/10.1075/ara1.21099.gre>.
- Hsu, W. 2014. "Measuring the Vocabulary Load of Engineering Textbooks for EFL Undergraduates." *English for Specific Purposes* 33: 54–65. <https://doi.org/10.1016/j.esp.2013.07.001>.
- Hultgren, A. K., C. Jensen, and S. Dimova. 2015. "Introduction: English-Medium Instruction in European Higher Education: From the North to the South." In *English-medium Instruction in European Higher Education*, edited by S. Dimova, A. K. Hultgren, and C. Jensen, 1–16. The Hague: De Gruyter Mouton. <http://www.degruyter.com/view/product/247850>.
- Jensen, C., and J. Thøgersen. 2011. "Danish University Lecturers' Attitudes Towards English as the Medium of Instruction." *Ibérica* 22: 13–33.
- Kling, J. 2015. "You Try with a Little Humor and you Just Get on with it": Danish Lecturers' Reflections on English-Medium Instruction." In *English-Medium Instruction in European Higher Education*, edited by S. Dimova, A. K. Hultgren, and C. Jensen, 201–222. De Gruyter Mouton.
- Kling, J., and L. S. Stæhr. 2012. The Development of the Test of Oral English Proficiency for Academic Staff (TOEPAS) (p. 28) [Technical Report]. Copenhagen: Centre for Internationalisation and Parallel Language Use. [http://cip.ku.dk/forskning/cip\\_publicationer/CIP\\_TOEPAS\\_Technical\\_Report.pdf/](http://cip.ku.dk/forskning/cip_publicationer/CIP_TOEPAS_Technical_Report.pdf/).
- Lasagabaster, D. 2022. "Teacher Preparedness for English-Medium Instruction." *Journal of English-Medium Instruction* 1: 48–64. <https://doi.org/10.1075/jemi.21011.las>.
- Lasagabaster, D., and A. Doiz. 2021. *Language Use in English-Medium Instruction at University: International Perspectives on Teacher Practice*. Abingdon: Routledge.
- Laufer, B. 2021. "Lexical Thresholds and Alleged Threats to Validity: A Storm in a Teacup?" *Reading in a Foreign Language* 33: 238–246.
- Laufer, B., and P. Nation. 1999. "A Vocabulary-Size Test of Controlled Productive Ability." *Language Testing* 16: 33–51. <https://doi.org/10.1177/026553229901600103>.



- Laufer, B., and G. Ravenhorst-Kalovski. 2010. "Lexical Threshold Revisited: Lexical Text Coverage, Learners' Vocabulary Size and Reading Comprehension." *Reading in a Foreign Language* 22: 15–30.
- Macaro, E. 2018. *English Medium Instruction*. Oxford: Oxford University Press.
- Macaro, E., S. Curle, J. Pun, J. An, and J. Dearden. 2018. "A Systematic Review of English Medium Instruction in Higher Education." *Language Teaching* 51: 36–76. <https://doi.org/10.1017/S0261444817000350>.
- Margić, B. D., and I. Vodopija-Krstanović. 2017. *Uncovering English-Medium Instruction: Global Issues in Higher Education*. Oxford: Peter Lang UK. <https://www.peterlang.com/view/title/61684>.
- McLean, S. 2021. "The Coverage Comprehension Model, its Importance to Pedagogy and Research, and Threats to the Validity with which it is Operationalized." *Reading in a Foreign Language* 33: 126–140.
- Milton, J. 2009. *Measuring Second Language Vocabulary Acquisition*. Bristol: Multilingual Matters.
- Milton, J. 2013. "Measuring the Contribution of Vocabulary Knowledge to Proficiency in the Four Skills." In *L2 Vocabulary Acquisition, Knowledge and Use: New Perspectives on Assessment and Corpus Analysis*, edited by C. Bardel, C. Lindqvist, and B. Laufer, 57–78. European Second Language Association.
- Molino, A., S. Dimova, J. Kling, and S. Larsen. 2023. *The Evolution of EMI Research in European Higher Education*. Abingdon: Routledge.
- Nation, P. 1983. "Testing and Teaching Vocabulary." *Guidelines* 5: 12–25.
- Pecorari, D., and H. Malmström. 2018. "At the Crossroads of TESOL and English Medium Instruction." *Tesol Quarterly* 52: 497–515.
- Pecorari, D., and H. Malmström. 2019. "Measuring Academic Vocabulary Knowledge: Triangulating Measures of Receptive and Productive Knowledge." Paper presented at the 54th RELC International Conference and 5th Asia-Pacific LSP and Professional Communication Association Conference, Singapore, 11–13 March, 2019.
- Pecorari, D., P. Shaw, and H. Malmström. 2019. "Developing a New Academic Vocabulary Test." *Journal of English for Academic Purposes* 39: 59–71. <https://doi.org/10.1016/j.jeap.2019.02.004>.
- Rose, H., S. Curle, I. Aizawa, and G. Thompson. 2020. "What Drives Success in English Medium Taught Courses? The Interplay between Language Proficiency, Academic Skills, and Motivation." *Studies in Higher Education* 45 (11): 2149–2161. <https://doi.org/10.1080/03075079.2019.1590690>.
- Schmitt, N., D. Schmitt, and C. Clapham. 2001. "Developing and Exploring the Behaviour of Two New Versions of the Vocabulary Levels Test." *Language Testing* 18: 55–88. <https://doi.org/10.1177/026553220101800103>.
- Stæhr, L. S. 2008. "Vocabulary Size and the Skills of Listening, Reading and Writing." *Language Learning Journal* 36: 139–152. <https://doi.org/10.1080/09571730802389975>.
- Tange, H. 2010. "Caught in the Tower of Babel: University Lecturers' Experiences with Internationalisation." *Language and Intercultural Communication* 10: 137–149. <https://doi.org/10.1080/14708470903342138>.
- Van Parijs, P. 2021. "Englishization as Trap and Lifeline." In *The Englishization of Higher Education in Europe*, edited by R. Wilkinson, and R. Gabriëls, 355–368. Amsterdam: Amsterdam University Press.
- Van Splunder, F., C. Verguts, T. De Moor, and S. De Paepe. 2022. "The Interuniversity Test of Academic English (ITACE): Assessing Lecturers' English Proficiency in Flanders." *Journal of English-Medium Instruction* 1: 255–274. <https://doi.org/10.1075/jemi.21007.van>.
- Warnby, M., H. Malmström, and K. Yang Hansen. 2023. "Linking Scores from Two Written Receptive English Academic Vocabulary tests: The VLT-Ac and the AVT." *Language Testing* 40: 548–575. <https://doi.org/10.1177/02655322221145643>.
- Wächter, B., and F. Maiworm. 2014. *English-Taught Programmes in European Higher Education: The State of Play in 2014*. Bonn: Lemmens.
- Webb, S. 2021. "Research Investigating Lexical Coverage and Lexical Profiling: What We Know, What We Don't Know, and What Needs to Be Examined." *Reading in a Foreign Language* 33: 278–293.
- Webb, S., Y. Sasao, and O. Balance. 2017. "The updated Vocabulary Levels Test: Developing and Validating Two New Forms of the VLT." *ITL-International Journal of Applied Linguistics* 168: 33–69. <https://doi.org/10.1075/itl.168.1.02web>.
- Werther, C., L. Denver, C. Jensen, and I. M. Mees. 2014. "Using English as a Medium of Instruction at University Level in Denmark: The Lecturer's Perspective." *Journal of Multilingual and Multicultural Development* 35: 443–462. <https://doi.org/10.1080/01434632.2013.868901>.

## Appendices

### Appendix 1

#### Test specifics and results for full sample (n = 130)

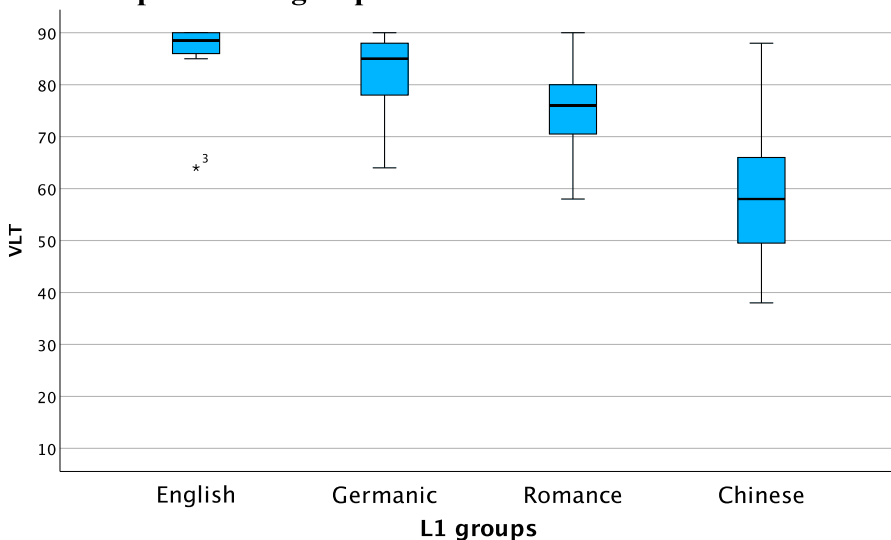
	Receptive vocabulary scoring					Productive vocabulary scoring				
	General vocabulary				Acad.	General vocabulary				Acad.
	3K	5K	10K	Tot VLT	AVT	3K	5K	10K	Tot PVLts	PAVts
N of items = max. possible score	30	30	30	90	57	18	18	18	54	52
Cronbach's <i>a</i>	.80	.86	.94	.95	.90	.84	.89	.90	.95	.96
Mean	28.58	27.24	20.10	75.92	47.38	11.85	11.02	8.68	31.55	25.96
Proportional mean score	.95	.91	.67	.84	.83	.66	.61	.48	.58	.50
Std. Error of Mean	0.21	0.32	0.7	1.13	0.67	0.35	0.4	0.44	1.4	1.16
Std. Deviation	2.39	3.65	7.98	12.86	7.65	4.03	4.60	5.06	12.99	13.23
Variance	5.72	13.33	63.64	165.33	58.5	16.23	21.13	25.63	168.81	175.12
Skewness	-2.36	-1.77	-.66	-1.12	-.85	-.54	-.29	-.06	-.28	-.08
Std. Error of Skewness	.21	.21	.21	.21	.21	.212	.212	.212	.212	.212
Kurtosis	6.07	2.86	-.47	.77	.19	-.57	-.99	-1.24	-1.14	-1.08
Std. Error of Kurtosis	.42	.42	.42	.42	.42	.422	.422	.422	.422	.422
Range	13	16	29	54	34	15	17	18	46	50
Minimum	17	14	1	36	23	1	1	0	5	1
Maximum	30	30	30	90	57	18	18	18	51	51
Percentiles										
10	25	23	8.1	57.1	37.10	6	4	1.10	14	8
25	28	25.75	15	68	42	9	7.75	4	20.75	14.75
50	30	29	21	79	49	13	11	9	33.50	26
75	30	30	28	87	54	15	15	13	44	36.25
90	30	30	29	89	56	17	17	15	47.90	42.90

### Appendix 2

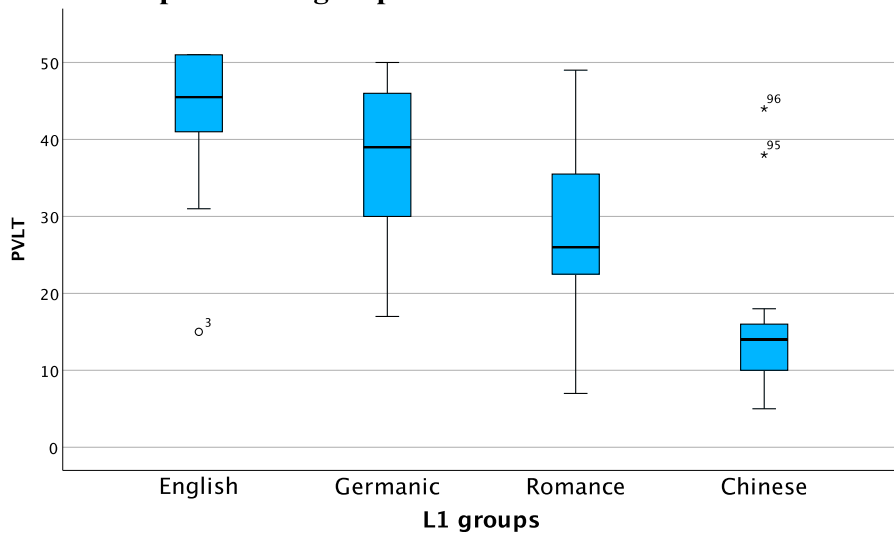
#### Boxplots of L1 groups

The boxplots in the figures below illustrate the differences between L1 groups. The large overlaps between certain groups indicate the non-significant findings reported, whereas the most substantial difference is found between the L1 Chinese group compared to the other L1 groups.

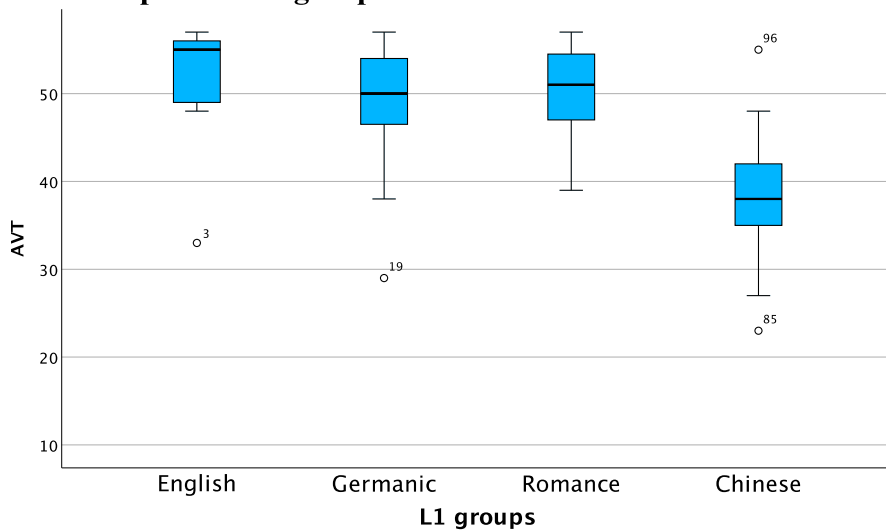
VLT – Boxplots of L1 groups



### PVLT – Boxplots of L1 groups



### AVT – Boxplots of L1 groups



### PAVT – Boxplots of L1 groups

