

## **Comparing EMI university reading materials with students' reading proficiency: Implications for admission testing**

Franz Holzknecht, Elisa Guggenbichler, Matthias Zehentner, Monique Yoder, Eva  
Konrad, Benjamin Kremmel

### **Abstract**

This study investigated to what extent the complexity of EMI university reading materials matches EMI students' reading proficiency. Text complexity and student proficiency were compared utilizing the Lexile<sup>®</sup> Framework for Reading text measures and reading scores based on reading test results, student questionnaire responses, and interviews with EMI lecturers. The results of the study indicate that, on average, Lexile text measures of EMI reading materials match students' Lexile reading scores well. However, the analyses also show that there are wide disparities between texts in terms of difficulty and students in terms proficiency. The questionnaire and interview data show that factors such as students' topical knowledge, text length, and text structure and organization are relevant aspects of perceived text complexity beyond word frequency and sentence length. In terms of assessment practices in EMI contexts, the findings of the study suggest that (1) EMI language admission tests should consider reading ability separately; (2) integrated reading/speaking and reading/writing tasks are authentic; (3) language admission tests in EMI settings should also consider text length; and (4) texts in admission tests should vary in terms of Lexile scores to reflect real-world EMI contexts.

### **Abstract (German)**

In dieser Studie wurde untersucht, inwieweit die Komplexität von Lesematerialien in EMI-Universitäten der Lesekompetenz der Studierenden entspricht. Textkomplexität und Lesekompetenz wurden anhand des Lexile<sup>®</sup>-Frameworks sowie auf Grundlage von Lesetestergebnissen, Antworten der Studierenden auf einen Fragebogen, und Interviews mit EMI-Dozenten verglichen. Die Ergebnisse der Studie zeigen, dass die Lexile-Werte der EMI-Lesematerialien im Durchschnitt gut mit den Lexile-Profilen der Studierenden übereinstimmen. Die Analysen zeigen jedoch auch, dass es große Unterschiede zwischen den Texten in Bezug

auf den Schwierigkeitsgrad und den Studierenden in Bezug auf die Lesefertigkeit gibt. Die Fragebogen- und Interviewdaten zeigen, dass neben der Worthäufigkeit und Satzlänge auch das thematische Wissen der Studierenden, die Textlänge, sowie die Textstruktur und -organisation eine Rolle für wahrgenommene Textkomplexität spielen. Im Hinblick auf die Beurteilungspraxis in EMI-Kontexten legen die Ergebnisse der Studie nahe, dass (1) EMI-Sprachzulassungstests die Lesefähigkeit separat berücksichtigen sollten; (2) integrierte Lese-/Sprech- und Lese-/Schreibaufgaben authentisch sind; (3) Sprachzulassungstests in EMI-Kontexten auch die Textlänge berücksichtigen sollten; und (4) Texte in EMI-Zulassungstests in Bezug auf Lexile-Werte variieren sollten, um reale EMI-Kontexte zu reflektieren.

### **Keywords**

EMI reading, assessing reading, Lexile framework, Aptis test

Holzkecht, F., Guggenbichler, E., Zehentner, M., Yoder, M., Konrad, E., & Kremmel, B. (2022). Comparing EMI university reading materials with students' reading proficiency: implications for admission testing. *Journal of English-Medium Instruction*, 1(2), 180–203. <https://doi.org/10.1075/jemi.21006.hol>

## **Background and literature review**

Teaching academic subjects through the use of the English language in non-English dominant contexts has seen a steady increase in popularity over the last decades (Dearden, 2014) and has become commonplace in universities around the globe. In Europe, this trend started in the Scandinavian countries and the Netherlands in the early 1950s, taking another 40 years to spread to Western and Eastern Europe by the 1990s (Coleman, 2006). The move towards internationalisation of higher education was expedited at the end of the last century by the introduction of the Bologna Declaration, aiming to promote the “international competitiveness of the European system of higher education” (European Ministers of Education, 1999). Additional momentum was gained through the European Union’s call to further language learning and multilingualism (European Commission, 2008).

Triggered by these significant changes in the education sector and the steady process of internationalisation, a growing number of European universities are now offering EMI programmes (see Wächter & Maiworm, 2014, for a comprehensive review). An increase of EMI degree programmes in higher education can also be observed outside Europe, for example in China (Galloway et al., 2017; Hu et al., 2014), Japan, Taiwan, Korea, or the Middle East (see Macaro et al., 2018 for an overview). In other parts of the world, studies on EMI teaching seem to be lacking. This is also the case for Africa, although English was established as the language of higher education in many African contexts in the early 2000s (Coleman, (2006). Our literature search revealed only one study, which found that EMI is widespread in South African and Namibian schools (2007). However, we were not able to find published empirical research on the use of EMI in other regions of Africa.

Additionally, comparative research of socioculturally and geographically diverse EMI contexts is rare (but see the study by Owen et al., 2021 discussed below). Assuming that local particularities would be prevalent in different educational contexts of EMI programmes, there is a need for more research into diverse EMI contexts. This is particularly the case for reading, where studies in EMI settings are scarce, as will be discussed in the following section.

### **The role of reading in EMI university settings**

Reading is perhaps the most fundamental academic skill, as it is a prerequisite for academic success in higher education (Hermida, 2009). Reading comprehension and vocabulary knowledge influence university grades (Van Lanen et al., 2000), and deficits in reading ability

can negatively impact proficiency in other subject areas (Schmeiser, 2009). Pecorari et al. (2012) found that textbooks and assigned readings are at the core of academic study and a helpful resource for content learning. It is thus not without reason that university students are said to “read for a degree” (Aldridge, 2019).

Well-developed reading skills are likewise crucial in L2 contexts for academic purposes within EMI instruction (Owen et al., 2021). L2 students of English, however, typically face additional challenges compared to L1 students: the difficulty of reading longer passages, disfluent reading, limited L2 background knowledge, the frequent use of integrated reading/writing tasks in academia, and insufficient language command in general (Grabe & Zhang, 2013). Reading of academic texts in L2 English takes more time compared to reading in an L1, and terminology is less easily retained in the L2 (Mežek, 2013).

Despite the rising number of EMI programmes as discussed above and the importance of reading for academic success, there is to date only scarce knowledge about reading proficiency and academic reading skills in EMI universities. Two studies on EMI reading were conducted in a Scandinavian context (e.g., Hellekjær, 2005, 2009; Mežek, 2013; Shaw & McMillion, 2008). Hellekjær (2005) investigated Norwegian students’ English L2 reading proficiency as they were transitioning from upper-secondary school to higher education. The study showed that two thirds of the students scored below Band 6 on the IELTS Academic Reading Module, i.e., below the threshold for admission to most British or Australian universities. In addition, about a third of university students experienced reading problems during their university studies (Hellekjær, 2009). Research by Shaw and MacMillion (2008), in contrast, showed that Swedish university students’ English reading performance was comparable to that of British L1 students; however the Swedish students were given extra time for task completion and relatively shorter texts.

More recently, Owen et al. (2021) compared reading demands of two EMI programmes in Sweden and Nepal in terms of suitability and (face) validity of the admission procedure. Using reading logs, student questionnaires, interviews, and TOEFL iBT test results, the authors identified differences in student proficiency, reading demands, and text materials between the two locations. However, more research is clearly necessary to investigate whether similar differences also apply to other contexts.

## Authenticity of EMI reading admission tests

While students who complete high school/upper-secondary school with appropriate prerequisites are automatically admitted to EMI programmes in many contexts, others may need to pass a standardized language test (Dimova, 2020). Such university entrance exams need to meet several test quality criteria to be valid instruments of measuring students' language competence and predicting their suitability for an EMI programme. Bachman and Palmer (1996) have argued that language tests for specific purposes, such as university entrance examinations, should strive to be as authentic as possible. In the case of university entrance tests, authentic language tests should reflect the critical features of relevant real-life academic activities by choosing content that future students will be likely to encounter in a non-test academic situation. Thus, the complexity of reading materials in admission tests should ideally mirror real-world EMI settings.

The last decade has seen drastic advances in computational linguistics, which have enabled fine-grained comparability analyses of reading materials. Corpus-based analyses allow for comparisons between the complexity of university reading materials and students' reading abilities, thereby operationalising reading ability as a latent variable in determining and predicting individuals' success at understanding reading materials (Stenner & Stone, 2004). Consequently, text analysis tools and measures of readability are useful indicators for test developers to target the intended authentic level of difficulty when aiming for predictive validity (Green et al., 2010).

One such tool is the Lexile Framework for Reading<sup>®</sup>, developed by MetaMetrics Inc. Similar to the majority of readability measures, the Lexile Framework includes semantic (i.e., word frequency) and syntactic (i.e., sentence length) factors to measure text complexity (Mesmer, 2008; Stenner et al., 2007). One advantage of the Lexile Framework is that it measures both text complexity and reader ability on the same scale through psychometrically linking standardized tests to the framework (Stenner, 2002). Lexile scores are measured in a common unit called Lexiles (L) and range from 0L for beginner level readers to over 2000L for very advanced readers in postgraduate university programmes. However, it is currently unknown how students' Lexile scores compare to reading requirements in non-English-dominant settings.

## Previous research on university reading materials using the Lexile Framework

Several studies have been conducted to compare the complexity of reading materials across different national and educational contexts utilizing the Lexile Framework. One of the earliest studies was carried out by Williamson (2004), who examined the complexity of high school and postsecondary reading materials in the American educational context. In this English-dominant setting, Williamson reported a monotonic increase in complexity from secondary education to military texts, and further to workplace and university reading material. Texts from the first two years of university study centred on a median of 1355L. In a later study, Williamson and colleagues (2016) found similar Lexile levels for reading materials used in universities across the United Kingdom.

Other studies identified lower Lexile levels for postsecondary reading. For example, Wilkins et al. (2010a) analysed textbooks used by entry-level US university students and reported Lexile scores similar to those found by Williamson et al. (2012) for materials used in grade 12 in the United States. These studies reported text measures of 1144L and 1130L, respectively. The Lexile scores reported by Wilkins et al.'s are thus comparable to the Lexile scores of materials from two EMI institutions included in this study (see Methodology section below). Finally, research by Koons et al. (2016) found slightly lower Lexile levels for textbooks used in Key Stage 4 (years 10 and 11) in the United Kingdom, with a median of 1030L.

All the above studies utilising the Lexile Framework have been conducted in English-dominant contexts; however, considering the increase in EMI programmes worldwide and the specific L2 factors relevant for reading described above, it seems crucial to also investigate the applicability of Lexile measures in L2 contexts. Only a small number of studies have investigated college readiness in terms of readability measures in EMI contexts in non-English dominant contexts. Williamson et al. (2016) report two studies embedded in an Asian context identifying a gap of 200L to 300L between students' reading abilities, secondary school textbooks, and postsecondary reading demands. However, to our knowledge, there is no research to date which has investigated student readiness in EMI university contexts in Europe and Africa utilizing the Lexile Framework. Given the increasing prevalence of EMI programmes in non-English dominant countries, and the important role of reading within these programmes, the present study aims to address this by investigating potential differences between reading demands of course texts and reading abilities of EMI students, and the ways in which course tutors perceive and mitigate such differences.

## Aim and research questions

In sum, studies into text complexity of university reading materials have hitherto been mostly confined to English-dominant contexts, e.g., the United States and the United Kingdom, paying little attention to the global and ever-growing trend of EMI. The main aim of the present study was thus to compare the textual demands of different EMI institutions in Europe and Africa (as expressed through Lexile text measures) with the reading ability of students at these institutions (as expressed through test scores and equivalent Lexile scores). Students' reading ability was assessed using Aptis Reading Tests, which have been formally linked to the Lexile Framework for Reading. The following research questions were addressed:

RQ1. To what extent do undergraduate and postgraduate EMI students' Aptis reading scores and equivalent Lexile scores match the demands of textual complexity that they will encounter in their diverse EMI university contexts?

RQ2. To what extent do undergraduate and postgraduate EMI students and lecturers from diverse disciplines perceive the texts in their EMI university contexts to be difficult?

## Methodology

The research questions were answered by four main data sources, collected across three different universities: (1) EMI students' Aptis Reading Test results with corresponding Lexile scores; (2) EMI students' questionnaire responses; (3) Lexile text measures of EMI university reading materials that the participating students were required to read as part of their studies; and (4) interviews with lecturers who used the reading materials with the participating students in class. The three participating universities were independent institutions in their respective countries rather than branch campuses of institutions in English dominant countries.

### Reading test and questionnaire

Students from three universities completed the Aptis Reading Test and filled in a questionnaire. The participants were 107 students from undergraduate EMI programmes in Egypt (n=68; 63.6%) and Lithuania (n=26; 24.3%), and a postgraduate EMI programme in Austria (n=13; 12.1%). The Egyptian and Lithuanian undergraduate students were either in the first (n=88), second (n=3), or third (n=3) year of their undergraduate degree. The Austrian candidates (n=13)

were all second-year Master's students. Across all three institutions, 61.6% of the students were female, and 37.9% were male; one student identified as non-binary. Most Egyptian and Lithuanian undergraduate students were younger than 20 years, with only two Egyptian students aged between 21 and 25 years. The Austrian postgraduate students' ages ranged between 21 to 25 years (61.5%) and 26 to 30 years (30.8%). One Austrian student was in their early thirties. To ensure anonymity, all students were assigned an individual reference number.

Overall, the students represented a mixed group of academic disciplines. The Egyptian sample contained students from a wide range of EMI subjects, such as engineering, international business administration, communication, or political science, among others. The Lithuanian candidates were relatively evenly distributed across different fields of the humanities and social sciences, whereas the Austrian postgraduate students were all majoring in ecology.

The participants included a considerable number of multilingual speakers. The test takers in the Egyptian sample reported to be L1 speakers of Arabic (n=68), English (n=12), French (n=5), and Turkish (n=1). The Lithuanian candidates included speakers of Ukrainian (n=9), Russian (n=8), Albanian (n=2), Lithuanian (n=2), Kazakh (n=2), Spanish (n=2), Afrikaans (n=1), Dari (n=1), English (n=1), and French (n=1). In comparison, German was the most frequently spoken language in the Austrian sample (n=8), followed by Italian (n=3), English (n=1), and Polish (n=1).

For organisational reasons, two different computer-based reading tests from the British Council were used in the study: the Aptis General Reading Test and the Aptis Advanced Reading Test. Both tests are language assessment tools for adults (16+) and they were developed for a variety of purposes such as "recruitment, workforce development or training" (British Council, 2020). The tests are linked to the Common European Framework of Reference for Languages (CEFR) and include a listening, reading, speaking, writing, and grammar/vocabulary component. In the current study only the reading components of the two tests were used. The Austrian postgraduate students completed the Aptis General Reading Test as their data was collected first and the Aptis Advanced Reading Test was not available to us at that time. The Egyptian and Lithuanian undergraduate students completed the Aptis Advanced Reading Test, which was the test made available to us by the British Council during this part of the data collection.

It should be noted that we used older versions of the tests, some of which have since been retired (in the case of Aptis General), as the British Council launched revised test versions in April 2020. These tests were not available to us at the time of data collection. The older



versions of the tests were developed using the same quality control procedures as the new versions, including professional item writing and item reviewing based on standardized test specifications, substantive field testing, and standard setting with official CEFR-linking. Despite this, we cannot rule out entirely that the results of the current study may be slightly different had the newer versions been used, or had the same version of the test been used by all participants.

While university admission testing is not listed as a specific purpose by the British Council, we chose to use Aptis tests as they have been formally linked to the Lexile Framework for Reading. In other words, students' scores on the Aptis tests can be translated into Lexile levels, which allowed us to directly compare text complexity of reading materials with students' reading ability on the same scale. Details about the linking process between Aptis and Lexile can unfortunately not be revealed for reasons of confidentiality.

After completing the Aptis test, students filled in an online questionnaire. The questionnaire included items about the university reading materials that students typically encountered during their studies. The first version of the questionnaire was piloted on five EMI university students, and the wording of two questions was changed slightly based on the pilot results.

#### Lexile analysis of reading materials

As a second step, reading materials from the participating EMI institutions were analysed in terms of Lexile text measures. The reading materials were sent to us by the participating EMI institutions. Before analysing the materials, we needed to ensure that the texts we used in our study were part of courses that all students attended. As the undergraduate students from Egypt and Lithuania were recruited from a wide range of EMI subject areas, we did therefore not use subject-area specific texts for the Lexile analysis, but texts from general core subjects which all students needed to complete (similar to the approach by Williamson, 2004). Successful completion of the core subjects was an integral part of all students' intended majors and therefore important for the students' overall academic success. For the Austrian sample, who were all majoring in ecology, we used subject-specific texts that all students had encountered during their studies (mostly peer-reviewed journal articles).

Table 1 provides an overview of the texts we analysed. As shown in the table, the texts varied widely with regards to number of words, particularly for the Egyptian sample. In terms of mean length, the Egyptian texts were the shortest on average ( $M=2,974$ ), followed by the

Lithuanian ( $M=4,274$  words) and Austrian texts ( $M=4,406$  words). On the other hand, the Egyptian text sample also included the longest text of all ( $Max=22,846$  words), which is considerably longer than the longest Austrian text ( $Max=8,152$ ).

**Table 1**

*Reading Materials Included in the Analysis for Each Institution*

Country of institution	N texts	N words per text			
		Mean	SD	Min	Max
Egypt	38	2,974	5,775	142	22,846
Lithuania	15	4,274	2,955	987	12,052
Austria	13	4,406	2,222	1,147	8,152

Interviews with lecturers

We also conducted semi-structured interviews (Riazi, 2016) with a total of seven lecturers of the participating students. The lecturers used the reading materials we analysed in their classes, so we gained insights into the texts' perceived level of complexity from the lecturer's perspective.

The interview participants were between 33 and 48 years old, with an average age of 42. Most lecturers ( $n=6$ ) had at least 14 years of experience in teaching at university, with one participant having taught for 3 years. In terms of subject areas, the lecturers from the EMI institution in Egypt were teaching the core subjects Scientific Thinking ( $n=2$ ) or Philosophical Thinking ( $n=1$ ), and the lecturers from Lithuania were teaching the core subjects History of Western Civilization ( $n=1$ ) and Intercultural Communication ( $n=1$ ). The Austrian interviewees were both lecturing various subjects in the postgraduate ecology course. The interviews with the lecturers were conducted online following standardized guidelines and lasted about 45 to 60 minutes.

Data analysis

The test and questionnaire data were analysed in three main steps. First, we received the detailed test results from the British Council and corresponding Lexile reading measures from MetaMetrics some weeks after each test administration. The data for all administrations were then combined. Second, the questionnaire data were combined and cleaned before linking them to the test results and Lexile score data through the candidate reference numbers. And third, to

answer RQ 1, the data were analysed separately for each participating institution using descriptive methods in SPSS (version 26 for Mac).

The interview data were transcribed and then analysed through holistic thematic analysis (Holliday, 2015). The data were coded using MaxQDA 2020 (VERBI software, 2019) following a coding scheme that was based on the interview questions. The interview contained six different thematic groups of questions displayed in Table 2. The data were coded according to these groups by one of the researchers, and the assigned codes were spot-checked by a second researcher. In all instances of spot-checking, the second researcher agreed with the assigned codes. The comments were then used to answer RQ 2. Selected comments are included in the results section.

**Table 2**

*Coding Scheme Based on Question Groups for the Interview Data*

code	description
(a) selection of reading materials	how the lecturers select reading materials for their courses; the types of reading materials they use; to what extent they select the materials themselves; which aspects of a text they consider when choosing materials; which role and purposes reading materials have in their teaching; whether they adapt texts themselves; or whether they take student feedback into account when selecting the texts
(b) reading tasks	what the students typically need to do with the texts in the courses; how the lecturers choose and assign the reading tasks; and how students generally react to the tasks
(c) student support	what lecturers do if they realise that students find it hard to comprehend the texts, and whether they think that their universities are offering enough support to students
(d) rating text complexity	lecturers were shown selected texts from the reading materials and were asked to order the texts according to complexity and to give reasons for their rank order
(e) students' reading test results*	whether the results of the study match the lecturers' teaching experience
(f) implications of the study	whether knowing about the results of the study has any implications on lecturers' work with reading materials

\*For this part of the interview we briefly outlined the students' test results and explained how they compared with the Lexile analysis of the reading materials used by the lecturers.

## Results and discussion

The first research question was to what extent EMI students' reading proficiency, as measured through the Aptis reading tests, compares to text complexity of authentic EMI reading materials. Table 3 below presents the candidates' results on the Aptis Reading Test, their corresponding CEFR levels as indicated by the British Council, and their Lexile reading measures, separately for the three EMI institutions. It is clear from the data that the Egyptian and Lithuanian reading measures share several key features. Compared to the Austrian postgraduate students, the Egyptian and Lithuanian undergraduate students achieved lower results overall. The Egyptian candidates had the lowest mean Aptis and Lexile results, scoring on average 30.9 on the Aptis Advanced Reading Test and 1128.6L according to the Lexile measures. Interestingly, however, the results ranged from a minimum of 16 to a maximum of 50 points on the Aptis test, which explains the relatively large standard deviation ( $SD=7.5$ ) and indicates important inter-individual differences. Similar observations can be made for the Lithuanian sample, where the standard deviation for candidates' Aptis scores was comparably large ( $SD=8.0$ ). On average, the Lithuanian candidates achieved slightly higher scores in both Aptis points and Lexile reading measures; nevertheless, none of the Lithuanian students attained the maximum of 50 points in the Aptis Advanced Reading Test. Correspondingly, the language proficiency levels according to CEFR scales range from B1 to C2 in the Egyptian and from B1 to C1 in the Lithuanian test taker group. By far, the highest average Aptis and Lexile reading measures were found for the Austrian postgraduate students. All Austrian candidates obtained between 40 and 50 points on the Aptis General Reading Test, with an average of 46.9 points and a relatively small standard deviation ( $SD=3.6$ ). Similarly, their Lexile reading measures ( $M=1420L$ ) were well above those of the other sample groups, and the Austrian participants' language proficiency levels were B2 or above, yet none had achieved a C2-level. Interestingly, the highest Lexile score was found in the Egyptian group with 1705L, which is well above the highest-scoring Austrian participant ( $Max=1610L$ ). The authors need to acknowledge, however, that there might be a ceiling effect in the Austrian sample, whose reading skills were measured based on the Aptis General Reading Test. As the Aptis General Reading Test is not intended to distinguish between the C levels, subsuming both levels as "C1 or above," and some Austrian students attained the maximum of 50 points, we can only state that they are at C level without, however, making statements about their full potential in reading. This is a caveat to comparing the test scores across institutions.

**Table 3***Participants' Aptis Reading Test Scores and Corresponding CEFR Levels and Lexile Reading Measures*

Country of institution	Measure	Aptis score*	CEFR level**	Lexile reading measure
Egypt (n=68)	M	30.9	3.9	1128.6L
	SD	7.5	0.8	168.5L
	Min	16	3	825L
	Max	50	6	1705L
Lithuania (n=26)	M	31.9	3.9	1130.8L
	SD	8.0	0.8	173.6L
	Min	18	3	865L
	Max	44	5	1445L
Austria (n=13)	M	46.9	4.8	1420L
	SD	3.6	0.4	208.7L
	Min	40	4	1090L
	Max	50	5	1610L

\*max. 50 points, participants in Egypt and Lithuania completed the Aptis Advanced Reading Test and participants in Austria completed the Aptis General Reading Test (see Methodology Section)

\*\*3=B1, 4=B2, 5=C1, 6=C2

The Lexile text measures of the reading materials are displayed in Table 4, again separately for each institution. Data obtained through the Lexile analyses of the texts reflect the results on candidates' scores outlined above. First, the analyses identified a gap between the Egyptian/Lithuanian undergraduate context and the Austrian postgraduate context when it comes to the reading materials the participants are assigned in their real-life university classes. The reading materials Lithuanian undergraduate students are confronted with had the lowest average level of lexical complexity ( $M=1119.3L$ ), followed by the Egyptian texts ( $M=1158.4L$ ). The reading materials assigned to the Austrian postgraduate students were markedly more challenging in terms of Lexile text measures ( $M=1412.3L$ ). Mann-Whitney Tests revealed statistically significant differences with large effect sizes between the Lexile text measures of the Egyptian and Austrian texts ( $p < .001$ ,  $Z=-4.5$ ,  $r=0.6$ ) and the Lithuanian and Austrian texts ( $p < .001$ ,  $Z=-3.3$ ,  $r=0.6$ ). Second, within the Egyptian and Lithuanian samples, text complexity measures varied widely. The least challenging text linguistically (Lexile measure of 790L), as well as the most challenging text (1660L), were both identified in the Egyptian sample. Similarly, the Lexile measures of the Lithuanian sample texts displayed the

largest standard deviation ( $SD=210.8L$ ) and ranged between 830L and 1460L. This indicates that the Egyptian and Lithuanian texts varied markedly in terms of text difficulty and linguistic demands as indicated through Lexile text measures. In contrast, the Austrian sample reading materials were much more homogeneous, ranging from 1260L to 1540L with a standard deviation of 89.3L.

**Table 4**

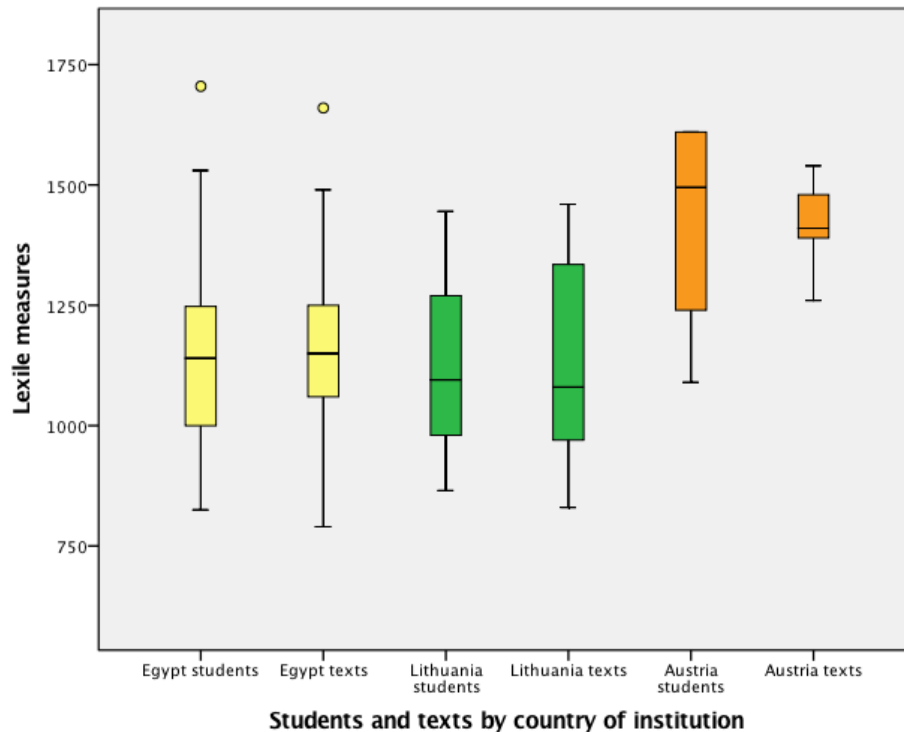
*Lexile text measures for each institution*

Country of institution	Measure	Lexile measure
Egypt (n=38)	M	1158.4
	SD	159.3
	Min	790
	Max	1660
Lithuania (n=15)	M	1119.3
	SD	210.8
	Min	830
	Max	1460
Austria (n=13)	M	1412.3
	SD	89.3
	Min	1260
	Max	1540

When comparing the Lexile text measures of the reading materials with the Lexile reading measures achieved by the student populations, the reading materials generally matched the target group proficiency (see Figure 1). Although the Egyptian students were presented with reading materials ( $M=1158.4L$ ) that were slightly above their average Lexile reading measures ( $M=1128.6L$ ), the difference was small. For the Lithuanian sample, the average Lexile score obtained by the students ( $M=1130.8L$ ) was slightly above the demands of the texts ( $M=1119.3L$ ). Likewise, the Austrian students achieved an average Lexile score of 1420L in the Aptis General Reading Test, which was slightly above the mean Lexile text measure identified in their course reading materials ( $M=1412.3L$ ).

**Figure 1**

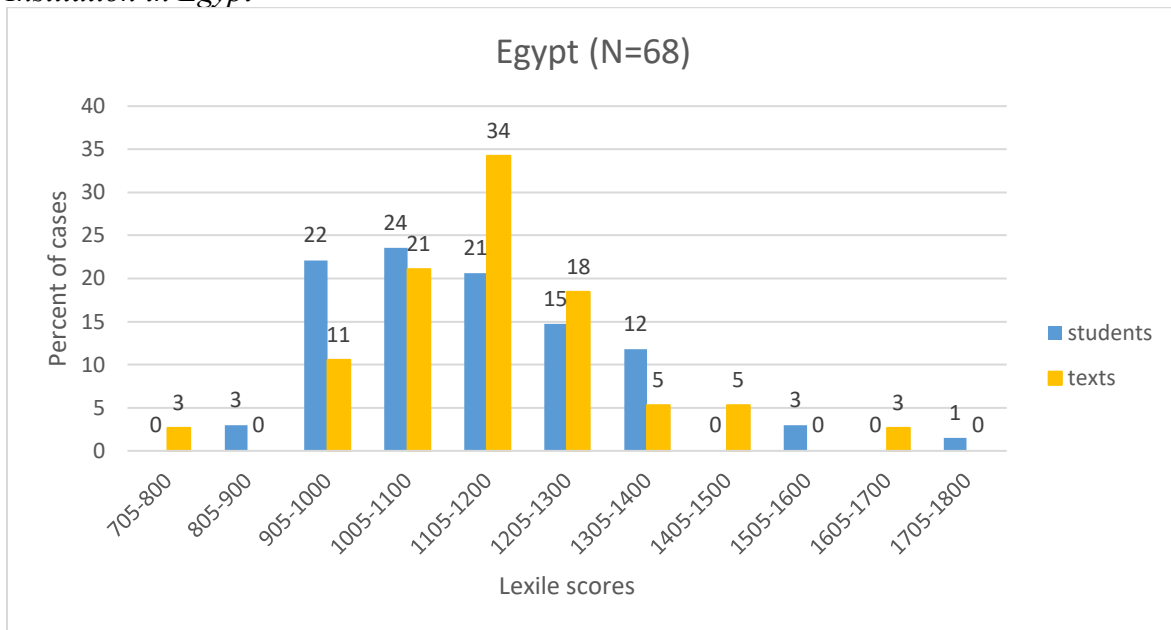
*Boxplot Comparison of Lexile Measures between Students and Texts across the Three Institutions*



However, the study also revealed important differences at an individual level, in that a considerable number of students would not be able to fully understand many texts, while other students would easily understand even the most complex texts (Figures 2-4). At the very least, many texts would fall outside the recommended difficulty margin by MetaMetrics, which is specified at 100L below or 50L above students' Lexile scores (MetaMetrics, 2022). For all three institutions, the Lexile measures of several texts were markedly above or below some of the students' Lexile measures. For example, in the Egyptian sample, about half of the students (49%) achieved Lexile reading measures below 1105L, but nearly two thirds of the texts (65%) displayed text difficulty measures above 1105L (Figure 2). Similarly, for the Lithuanian sample, 77 percent of all students obtained Lexile reading measures below 1305L, but a third of the texts (33%) were measured at or above 1305L (Figure 3). The Austrian sample, on the other hand, showed two extremes. While 23 percent of the Austrian students achieved Lexile reading measures below 1205L and thus lower levels than even the simplest text in the Austrian sample (measured at 1260L), 38 percent of students were placed at higher levels than the most difficult text (< 1600L, see Figure 4). In sum, while the average students' Lexile reading measures match the reading materials' average complexity relatively well, there are mismatches between individual students' proficiency and the reading materials' difficulty.

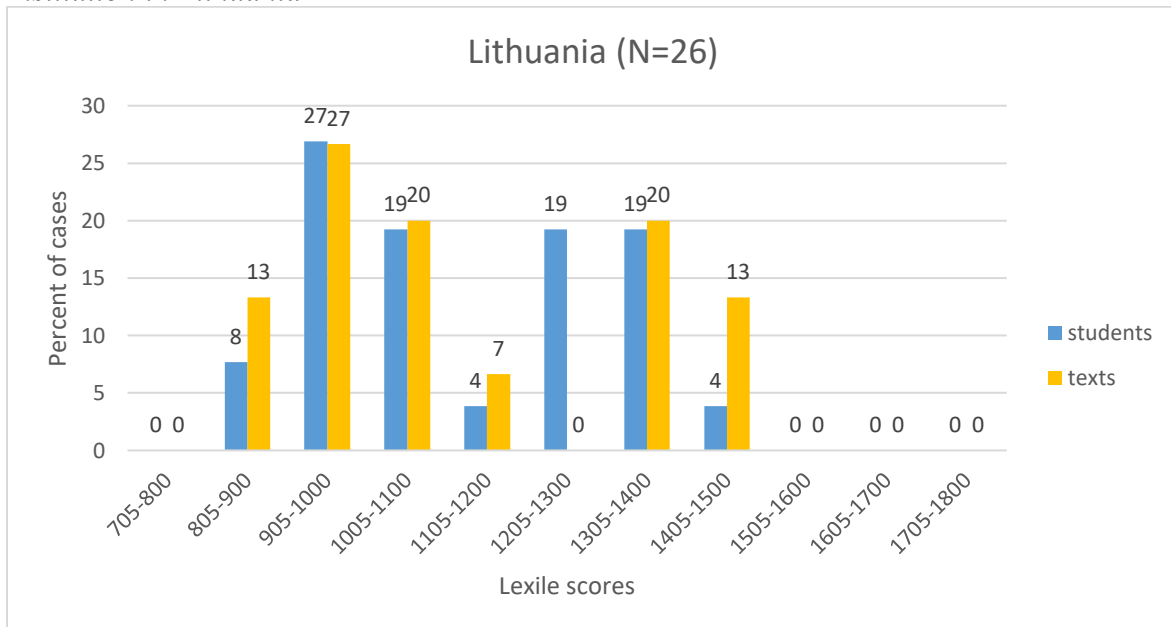
**Figure 2**

*Comparison of Lexile Measures between Students and Texts (% of Cases at Each Level) for the Institution in Egypt*



**Figure 3**

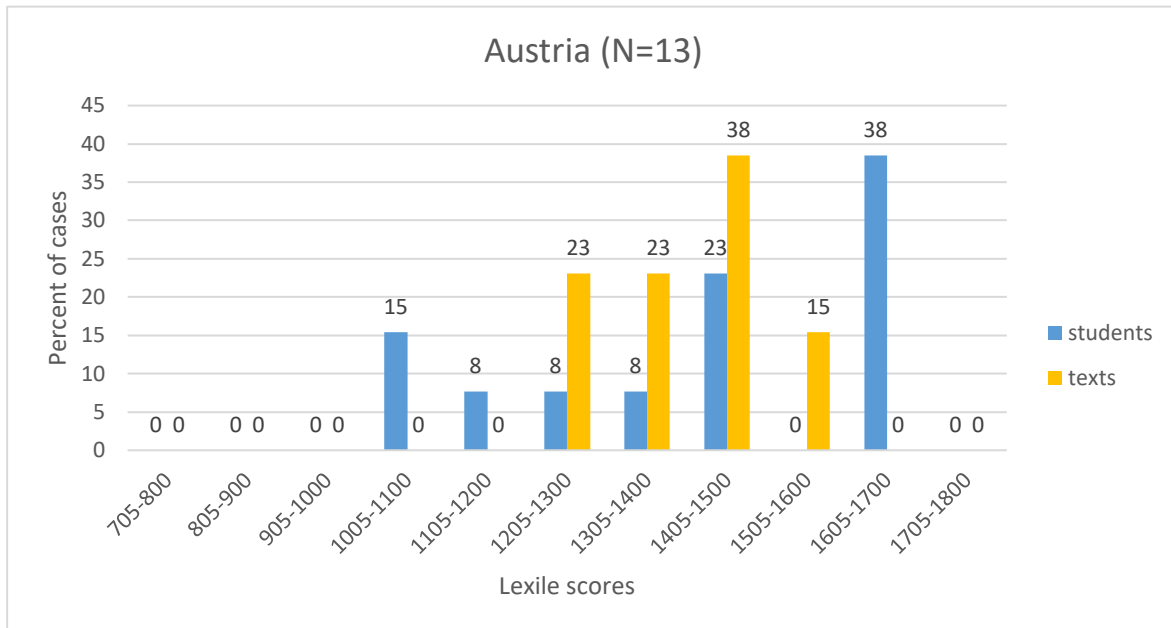
*Comparison of Lexile Measures between Students and Texts (% of Cases at Each Level) for the Institution in Lithuania*





**Figure 4**

*Comparison of Lexile Measures between Students and Texts (% of Cases at Each Level) for the Institution in Austria*



#### Implications for admission testing

These results raise the question of what level of reading should ideally be targeted by EMI admissions tests. If tests target the lowest level of text difficulty, many students will struggle with more complex texts they will encounter in real-world EMI settings. On the other hand, if tests target the bulk of the distribution, many students will fail to pass admissions criteria although they would be able to understand a substantial number of texts covered in the courses. One possible solution may be to include texts at different levels in admissions tests to reflect real-world EMI settings more accurately and to report results for different difficulty levels separately. Such tests would provide more detailed information on the match between students' reading proficiency and the courses' text difficulty ratings, which could potentially lead to fairer admissions decisions.

The interviews with lecturers revealed insights into why the complexity of reading materials might deviate from students' level of proficiency. First, the data indicate that several lecturers struggled with estimating text complexity and generally must rely on their experience rather than objective external criteria. Lecturers were surprised that many students would have difficulty in understanding the course reading materials, as illustrated in the following example from the interview data:

Interviewer 1: So 50% would struggle to understand 2/3 of the texts which we analysed.

I06: That would seem high. Over my course.

Second, the observed range in student proficiency confirms findings by Lam and Maiworm (2014), who report that heterogeneity in students' English proficiency is perceived as a significant problem by many institutional coordinators and programme directors in EMI institutions (but note that Lam and Maiworm looked at general English proficiency and not at reading ability specifically). Although a certain range in proficiency may be desirable in EMI programmes so that students can support each other, it seems problematic that several students would not be able to fully understand the simplest texts in their courses. This was also raised as a concern in the interviews, where the lecturers mentioned that some students' English proficiency is insufficient for the demands of the degree programme, as illustrated by this example from a lecturer at the Egyptian institution:

I06: I do not know how they evaluate it [the students' reading proficiency]. I have been told that they are expected to be fluent in English, both reading and writing, and I definitely have students who struggle. (Erm) Not the majority of the students, but there I /each semester I have a couple students so I can tell they are not where they are expected to be language-wise, and in that case I have my TA [teaching assistant] help.

In the current study, students of two institutions had to show proof of English proficiency through TOEFL or IELTS scores to be admitted for study; the institution in Lithuania stipulated CEFR B2 as the minimum requirement and the institution in Egypt set their cutoff scores between B2 and C1. However, neither institution demanded minimum levels for reading, but only for overall scores across all skills or, in the case of the Egyptian institution, for overall scores and writing scores. Given the fact that several students in the current study demonstrated insufficient reading proficiency for the demands in their courses, assessing reading separately in admission testing for undergraduate programmes may thus be necessary.

The discrepancy between student ability and text difficulty in the Egyptian and Lithuanian programmes may also be related to the type of reading tested in the language admission tests, as TOEFL and IELTS tests may be to some extent incompatible with the actual

reading demands of the EMI programmes. This would mirror findings by Owen et al. (2021), who argue that international high-stakes language examinations can be useful tools to test students' general language proficiency, but might require localisation and adaptations to authentically measure academic reading demands of EMI programmes. However, this potential link cannot be untangled with the data from the current study.

For the Austrian postgraduate course, no entry requirements in terms of language ability were mentioned in the admission criteria. This is concerning, given that 23% of students in that programme would struggle to understand even the simplest texts. Screening incoming students in terms of English reading proficiency may thus also be necessary for EMI postgraduate programmes, particularly if students have no previous (undergraduate) experience of EMI instruction.

With regards to text complexity, the average Lexile text measures of EMI reading materials in our study are comparable to Lexile text measures found in previous research in an L1 context. Texts for beginner students of the EMI undergraduate programmes in Egypt and Lithuania centred around 1160L and 1120L respectively, which compares to Lexile text measures of beginner L1 university students of English in the United States (Wilkins et al., 2010, who report a median of 1144L), and to grade 11 and grade 12 texts in the United States (Williamson et al., 2012 who report median of 1130L). Koons et al. (2016) report a slightly lower median of 1030L for key stage 4 texts in the United Kingdom. In comparison, Williamson (2004) reports median Lexile text measures of 1355L for first and second year L1 U.S. university texts combined, thus indicating that text complexity levels rise throughout the first two years of university. Correspondingly, text complexity in the Austrian postgraduate programme was higher than in the undergraduate programmes, with an average of around 1420L. Thus, EMI postgraduate level reading materials were above the typical level of L1 postsecondary reading materials as identified by Smith and Williamson (2016) in their meta-analysis (1300L), as well as above levels of other readability studies in L1 (undergraduate) postsecondary reading (Williamson, 2004; Williamson et al., 2016).

Although average Lexile text measures aligned well with comparable L1 settings, we also found that texts used in EMI teaching varied widely with regards to complexity as measured through Lexile text measures. This was more pronounced for EMI undergraduate courses, where texts ranged between 790L and 1660L. The range of text complexity levels identified in the current study is thus greater than comparable studies in an English-dominant university context, where reading materials are generally more homogeneous in terms of text complexity (Wilkins et al., 2010b; Williamson, 2004; Williamson et al., 2016). As argued

above, in terms of admission testing in EMI contexts, these findings suggest that local tests of reading should include texts at a variety of levels (as, for example, based on Lexile measures) to better reflect real-world EMI contexts.

What makes texts difficult?

To answer this research question, we drew on student questionnaire responses and the interviews with lecturers. Most students across the three institutions stated that they found the texts “easy” or “rather easy,” while only a smaller number (n=24) reported that they found the texts “rather difficult.” Reasons given for perceived text difficulty were “unfamiliar words and phrases” (n=14), “unfamiliar topics and concepts” (n=9), “structure and organization of texts” (n=6), “length of texts” (n=6), and “unfamiliar grammatical structures” (n=3).

Similar reasons emerged from the qualitative analysis of the interviews conducted with lecturers at each of the EMI institutions. When asked what they considered when estimating text complexity, all lecturers (n=7) mentioned evaluating linguistic aspects. One linguistic aspect mentioned by the lecturers was vocabulary, referring both to the level of formality and the amount of subject-specific terminology. Vocabulary size is generally acknowledged to be of critical importance for L2 reading comprehension, with lexical difficulty accounting for approximately 80% of the variance in many readability studies (Alderson, 2000). Our findings thus mirror previous research in this regard and corroborate vocabulary frequency as one of the main underlying factors in calculating readability scores such as the Lexile Framework.

The second factor influencing scores on the Lexile Framework is sentence length. Five out of seven lecturers reported that aspects of morphosyntax, in particular grammatical structures and sentence length, contributed to their judgement of increased text complexity. Students themselves, however, did not directly mention long sentences as a factor for text complexity (but only long texts), perhaps because it was not included as a pre-defined category in the student questionnaire.

Our study identified three additional factors which play a role in determining a text’s perceived readability. One of them is the students’ familiarity with topics and concepts, which was the second most important factor for text complexity reported by students. Parallel to Owen et al.’s (2021) participants, students in our sample felt that topic familiarity affected their performance. Likewise, five out of seven lecturers reported that conceptual difficulty was a factor they considered when judging text complexity. Lecturers also indicated that they often assign integrated reading/writing or reading/speaking tasks to introduce new topics. Based on

these findings, it seems important to refer to DuBay (2004), who argued that readability indices should also consider contextual factors, such as background knowledge and subject-specific knowledge, when calculating text complexity scores.

Another factor contributing to text difficulty mentioned by both students and lecturers is text length, in that longer texts were more difficult to comprehend than shorter texts. Text length is also a key feature in many language proficiency frameworks such as the CEFR, where the comprehension of “lengthy, complex texts” is only expected at C1 level, while comprehending “short, simple texts” can be achieved at A2 (Council of Europe, 2018, p. 60). These findings suggest that language admission tests in EMI settings should thus also consider text length in their test specifications.

Finally, another variable which seems to influence text complexity is the texts’ structure and organization, as undergraduate students indicated this several times in their questionnaire responses. Lecturers also considered the appropriate use of cohesive devices and contextual markers to foster text comprehension for students. Three lecturers from the natural sciences further attributed text complexity to the writing styles of specific scientific journals and genres, which are dense in information. While it may be challenging for a framework like Lexile to incorporate topic familiarity, incorporating indices of text coherence and cohesion might be feasible and promising for future revisions of the Lexile Framework, as has also been highlighted in research by Crossley and colleagues (2007, 2008, 2011, 2017).

## **Limitations**

Even though this study has furthered our understanding of the need for increased focus on reading admission testing in EMI contexts, the generalisability of the findings is subject to several limitations. First, the scope of this study is limited by a relatively small sample size, especially with regards to the individual student groups, particularly the Lithuanian and Austrian groups. However, including students from different EMI contexts also allowed us to compare reading demands across institutions. In addition, it seems important to include currently underrepresented geographical regions in research on EMI.

A related limitation lies in the small sample size and subjective nature of the questionnaire and the interviews with lecturers (see also Seliger & Shohamy, 2001). Although the interviews added to our understanding of the lecturers’ rationale behind selecting reading materials and tasks, the findings should be interpreted with caution, as the informants came from various subject-specific backgrounds and differed in their teaching experiences. In

addition, the lecturers' attitudes to difficulty levels may have been influenced by material they had used in previous courses or degree programmes. Similarly, students' perceptions were measured solely by means of a questionnaire; however, additional follow-up interviews may have shed more light on the extent to which students found the texts difficult, and the reasons thereof. Thus, the data offer preliminary insights, rather than a comprehensive overview, of what aspects students and lecturers consider when judging the level of difficulty of various text sources.

Due to our research design and participant sampling, it was important that we only included texts that all students across an institution encountered in their studies. As the Egyptian and Lithuanian students were sampled from a large variety of academic disciplines, we therefore chose texts from their majors' core subjects, rather than subject-specific texts. Although these core subjects were an integral part of all students' majors, the results of the study may be slightly different had we used texts specific to the students' academic disciplines.

## **Conclusion**

This study investigated to what extent the complexity of EMI university reading materials matches EMI students' reading proficiency. Text complexity and student proficiency were compared utilizing the Lexile<sup>®</sup> Framework and reading scores based on Aptis test results, student questionnaire responses, and interviews with EMI lecturers. In terms of assessment practices in EMI contexts, the findings of the study suggest that (1) EMI language admission tests should ideally consider reading ability separately, rather than as part of overall language proficiency; (2) integrated reading/speaking and reading/writing tasks would be authentic, as many lecturers assign these kinds of tasks in their courses; (3) language admission tests in EMI settings should also consider text length in their test specifications; and (4) diagnostic and/or admission tests of reading should not only include texts at a particular level (as based on Lexile measures), but texts should vary in terms of Lexile scores to reflect real-world EMI contexts.

## **References**

- Alderson, J. C. (2000). *Assessing reading*. Cambridge University Press.
- Aldridge, D. (2019). Reading, engagement and higher education. *Higher Education Research & Development*, 38(1), 38–50. <https://doi.org/10.1080/07294360.2018.1534804>
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford University Press.

- British Council. (2020). *Aptis*. <https://www.britishcouncil.ge/en/exam/aptis/what>
- Coleman, J. A. (2006). English-medium teaching in European higher education. *Language Teaching*, 39(1), 1–14. <https://doi.org/10.1017/S026144480600320X>
- Crossley, S. A., Allen, D. A., & McNamara, D. S. (2011). Text readability and intuitive simplification: A comparison of readability formulas. *Reading in a Foreign Language*, 23(1), 84–101.
- Crossley, S. A., Greenfield, J., & McNamara, D. S. (2008). Assessing text readability using cognitively based indices. *TESOL Quarterly*, 42(3), 475–493. <https://doi.org/10.1002/j.1545-7249.2008.tb00142.x>
- Crossley, S. A., Louwse, M., McCarthy, P., & McNamara, D. (2007). A linguistic analysis of simplified and authentic texts. *The Modern Language Journal*, 91, 15–30. <https://doi.org/10.1111/j.1540-4781.2007.00507.x>
- Crossley, S. A., Skalicky, S., Dascalu, M., McNamara, D. S., & Kyle, K. (2017). Predicting text comprehension, processing, and familiarity in adult readers: New approaches to readability formulas. *Discourse Processes*, 54(5–6), 340–359. <https://doi.org/10.1080/0163853X.2017.1296264>
- Dearden, J. (2014). English as a medium of instruction: A growing global phenomenon. In *English language and higher education*.
- Dimova, S. (2020). English language requirements for enrolment in EMI programs in higher education: A European case. *Journal of English for Academic Purposes*, 47, 100896. <https://doi.org/10.1016/j.jeap.2020.100896>
- DuBay, W. (2004). *Principles of readability*. Impact Information. <http://www.impact-information.com/impactinfo/readability02.pdf>
- European Commission. (2008). Multilingualism: An asset for Europe and a shared commitment. In *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee*. European Commission.

- European Ministers of Education. (1999). *The Bologna Declaration of 19 June 1999* (Issue June). The European Higher Education Area. <https://doi.org/10.1007/978-3-319-20877-0>
- Ferencz, I., Maiworm, F., & Mitic, M. (2014). Part II – Traits and daily operation of ETPs. In B. Wächter & F. Maiworm (Eds.), *English-taught programmes in European higher education: The state of play in 2014* (pp. 63–97). Lemmens Medien.
- Galloway, N., Kriukow, J., & Numajiri, T. (2017). Internationalisation, higher education and the growing demand for English: An investigation into the English medium of instruction (EMI) movement in China and Japan. In *British Council ELT Research Papers* (Vol. 17, Issue 2).
- Grabe, W., & Zhang, C. (2013). Reading and writing together: A critical component of English for academic purposes teaching and learning. *TESOL Journal*, 4(1), 9–24. <https://doi.org/10.1002/tesj.65>
- Green, A., Ünalı, A., & Weir, C. (2010). Empiricism versus connoisseurship: Establishing the appropriacy of texts in tests of academic reading. *Language Testing*, 27(2), 191–211. <https://doi.org/10.1177/0265532209349471>
- Hellekjær, G. O. (2005). *The acid test: Does upper secondary EFL instruction effectively prepare Norwegian students for the reading of English textbooks at colleges and universities? [Unpublished PhD thesis]*. University of Oslo, Norway.
- Hellekjær, G. O. (2009). Academic English reading proficiency at the university level: A Norwegian case study. *Reading in a Foreign Language*, 21(2).
- Hermida, J. (2009). The importance of teaching academic reading skills in first-year university courses. *The International Journal of Research and Review*, 20–30.
- Holzknicht, F., Guggenbichler, E., Zehentner, M., Yoder, M., Konrad, E., & Kremmel, B. (2022). Comparing EMI university reading materials with students' reading



- proficiency: Implications for admission testing. *Journal of English-Medium Instruction*, 1(2), 180–203. <https://doi.org/10.1075/jemi.21006.hol>
- Hu, G., Li, L., & Lei, J. (2014). English-medium instruction at a Chinese university: Rhetoric and reality. *Language Policy*, 13(1), 21–40. <https://doi.org/10.1007/s10993-013-9298-3>
- Koons, H., Bush, L., & Sanford-Moore, E. E. (2016). *Primary and secondary textbook complexity in England*.
- Lam, Q. K. H., & Maiworm, F. (2014). Part III - English in the classroom and beyond. In B. Wächter & F. Maiworm (Eds.), *English-taught programmes in European higher education: The state of play in 2014* (pp. 99–118). Lemmens Medien.
- Macaro, E., Curle, S., Pun, J., An, J., & Dearden, J. (2018). A systematic review of English medium instruction in higher education. *Language Teaching*, 51(1), 36–76. <https://doi.org/10.1017/S0261444817000350>
- Mesmer, H. A. E. (2008). *Tools for matching readers to texts: Research-based practices*. Guildford Press.
- Mezek, S. (2013). *Advanced second-language reading and vocabulary learning in the parallel-language university*. Stockholm University.
- Mežek, Š. (2013). Multilingual reading proficiency in an emerging parallel-language environment. *Journal of English for Academic Purposes*, 12(3), 166–179. <https://doi.org/10.1016/j.jeap.2013.02.001>
- Owen, N., Shrestha, P. N., & Hultgren, A. K. (2021). Researching academic reading in two contrasting English as a medium of instruction contexts at a university level. In *ETS Research Report Series*. John Wiley and Sons Inc. <https://doi.org/10.1002/ETS2.12317>
- Pecorari, D., Shaw, P., Irvine, A., Malmström, H., & Mežek, Š. (2012). Reading in tertiary education: Undergraduate student practices and attitudes.

[Http://Dx.Doi.Org/10.1080/13538322.2012.706464](http://dx.doi.org/10.1080/13538322.2012.706464), 18(2), 235–256.

<https://doi.org/10.1080/13538322.2012.706464>

Riazi, A. M. (2016). *The Routledge encyclopedia of research methods in applied linguistics*.

Routledge.

Schmeiser, C. B. (2009). *Impact of literacy on college and career readiness and STEM success*.

Seliger, H. W., & Shohamy, E. (2001). *Second language research methods* (6th ed.). Oxford University Press.

Shaw, P., & McMillion, A. (2008). Proficiency Effects and Compensation in Advanced Second-Language Reading. *Nordic Journal of English Studies*, 7(3), 123.

<https://doi.org/10.35360/NJES.104/METRICS/>

Smith, M., & Williamson, G. L. (2016). Aligning the K-12 academic journey with a postsecondary destination. *The State Education Standard*, 16(1), 14–18.

Stenner, A. J. (2002). Matching students to text: The targeted reader (Scholastic professional paper). *Scholastic Red*, 1–7.

Stenner, A. J., Sanford-Moore, E. E., & Burdick, D. (2007). *The Lexile Framework® for Reading: Technical report* (Issue April).

Stenner, A. J., & Stone, M. H. (2004). Does the reader comprehend the text because the reader is able or because the text is easy? *International Reading Association*.

Uys, M., Walt, J. Van Der, Berg, R. Van Den, & Botha, S. (2007). *English medium of instruction: A situation analysis*. 27(1), 69–82.

Van Lanen, R. J., Lockie, N. M., & McGannon, T. (2000). Predictors of nursing students' performance in a one-semester organic and biochemistry course. *Chemical Education Research*, 77(6), 767–770. <https://doi.org/10.1021/ed077p767>

- Wilkins, C., Hartman, J., Howland, N., & Sharma, N. (2010a). How prepared are students for college level reading? Applying a Lexile®-based approach. *Issues & Answers Report, REL 2010–No. 094, 0*.
- Wilkins, C., Hartman, J., Howland, N., & Sharma, N. (2010b). How prepared are students for college level reading? Applying a Lexile®-based approach. In *Issues & Answers Report*.
- Williamson, G. L. (2004). 1350L Lexile: Matching readers to texts: Student readiness for postsecondary options. *The Lexile Framework for Reading*, 1–30.
- Williamson, G. L., Koons, H., Sandvik, T., & Sanford-Moore, E. E. (2012). *The text complexity continuum in grades 1-12*.
- Williamson, G. L., Sandvik, T., Stenner, A. J., & Johnson, A. (2016). *Complexity of university texts in the United Kingdom*. 0–3.