

Research into Reading Grants

ISSN 2755-2543 Report 2/2022



Comparing EMI university reading materials with students' reading proficiency utilizing Lexile® measures, Aptis test results, student questionnaire responses, and interviews with lecturers

RESEARCH IN EFL READING AND READING ASSESSMENT

Franz Holzknicht, Language Testing Research Group Innsbruck, University of Innsbruck, Austria
(Current affiliation: University of Teacher Education in Special Needs, Zurich, Switzerland)

Elisa Guggenbichler, Language Testing Research Group Innsbruck, University of Innsbruck, Austria

Matthias Zehentner, Language Testing Research Group Innsbruck, University of Innsbruck, Austria

Monique Yoder, Michigan State University, USA and LCC International University, Lithuania

Eva Konrad, Language Testing Research Group Innsbruck, University of Innsbruck, Austria

Benjamin Kremmel, Language Testing Research Group Innsbruck, University of Innsbruck, Austria

ABSTRACT

This study investigated to what extent the complexity of English as a medium of instruction (EMI) university reading materials matches EMI students' reading proficiency. Text complexity and student proficiency were compared utilizing Lexile® Framework for Reading text measures and reading measures based on test results, questionnaire responses, and interviews. Specifically, the study combined data from 107 students' Aptis reading test scores and equivalent Lexile reading measures with student questionnaire responses, Lexile text measures of 66 authentic texts used in EMI university teaching, and interviews with 7 lecturers. The participants were recruited from three different EMI institutions in Europe (Austria and Lithuania) and Africa (Egypt).

The results of the study indicate that, on average, Lexile text measures of EMI reading materials match students' Lexile reading measures relatively well and text complexity is similar to comparable L1 university settings. However, the analyses also showed that there were quite wide disparities between: 1) students in terms of proficiency; and 2) between texts in terms of difficulty, as indicated through Lexile reading measures and text measures. While some students displayed Lexile reading measures that were above the text measures of all the texts used in EMI teaching, a substantial number of students' Lexile reading measures was below the level of even the least complex texts.

The second major finding of the study is that readability indices such as the Lexile text measure, while being useful indicators for the general difficulty of reading materials, appear to only provide a relatively narrow view of text complexity. The questionnaire and interview data showed that factors such as students' topical knowledge, text length, text structure and organization, and reading tasks, as well as individual differences between students, were relevant aspects of perceived text complexity beyond word frequency and sentence length.

AUTHORS

Franz Holzknecht

Franz Holzknecht is postdoctoral researcher at the University of Teacher Education in Special Needs Zurich, where he specializes in language testing, particularly in sign language assessment and L2 listening assessment. Franz has published in the journals *Language Testing*, *Assessing Writing*, *Papers in Language Testing and Assessment*, and *The International Journal of Listening*, and in edited volumes such as the *Routledge Handbook of Language Testing*. He is a recipient of the UKALTA Postgraduate Presentation Award and the Robert Lado Memorial Award. The research presented in this report was conducted while he was working for the Language Testing Research Group Innsbruck.

Elisa Guggenbichler

Elisa Guggenbichler studied French and English as foreign languages at the University of Innsbruck, Austria, where she gained an MA degree in Teaching and a BA degree in Translation (German/French/English). She is currently a member of the Language Testing Research Group Innsbruck and a PhD candidate in Education, focusing on language assessment. Elisa's interests lie in assessing foreign languages (especially English and French), task development, cross-linguistic mediation, and diagnostic assessment of listening.

Matthias Zehentner

Matthias Zehentner holds an MA degree in Teaching English and Mathematics from the University of Innsbruck, and an MA in Language Testing from Lancaster University. He teaches English and Mathematics at a secondary vocational school in Innsbruck and was employed as a researcher and lecturer at the University of Innsbruck from 2009 to 2020. His research interests include Mathematics assessment, L2 writing and speaking assessment, automated assessment, and classroom assessment. He is currently involved in the development of the Austrian Mathematics secondary school-leaving exam.

Monique Yoder

Monique Yoder is a PhD student in the Second Language Studies Program at Michigan State University (USA) and an MA TESOL affiliate faculty member with LCC International University (Lithuania). Within language assessment, she researches topics on test-taker responses, stakeholder language assessment literacy, young learner assessment, English for academic purposes writing assessment, and test development in contexts with limited resources. Within second language acquisition, she researches L2 writing pedagogy and support within higher education.

Eva Konrad

Eva Konrad is a researcher at the Faculty of Teacher Education at the University of Innsbruck and a member of the Language Testing Research Group Innsbruck. She holds an MA in English and American Studies from the University of Innsbruck and an MA in Language Testing from Lancaster University. Her research interests are assessment literacy, L2 writing assessment, the effect of dictionaries on writing, item writer training, and the development, evaluation, and quality control of reading and listening items. She conducts language testing workshops for schools and other educational institutions.

Benjamin Kremmel

Benjamin Kremmel is Head of the Language Testing Research Group Innsbruck (LTRGI) at the University of Innsbruck, Austria, where he teaches and researches language learning, teaching, and assessment. His research interests include vocabulary assessment, L2 reading assessment, diagnostic language testing, language assessment literacy, and SLA. His work has been published in *Language Testing*, *Language Assessment Quarterly*, *Applied Linguistics*, *Language Teaching*, *Studies in Second Language Acquisition*, and *TESOL Quarterly*. He currently serves as Book Review Editor for *Language Testing* and is the Associate Editor for the Assessment and Testing section of the *Encyclopedia of Applied Linguistics* (Wiley).

CONTENTS

1. BACKGROUND AND LITERATURE REVIEW	7
1.1 The role of English as a medium of instruction (EMI) in international academia	7
1.1.1 Reasons for offering EMI courses and associated challenges	8
1.1.2 Subject areas and admission criteria for EMI courses	9
1.2 The role of reading in EMI university settings	10
1.3 Testing reading skills as part of EMI university admission tests	11
1.4 Measuring text complexity using the Lexile Framework® for Reading	13
1.5 Previous research on university reading materials using the Lexile Framework®	15
2. RESEARCH QUESTIONS	17
3. METHODOLOGY	18
3.1 Reading test and questionnaire	18
3.1.1 Aptis reading tests	18
3.1.2 Questionnaire	19
3.1.3 Reading test participant recruitment	20
3.1.4 Final reading test participant sample	23
3.1.5 Test administration	27
3.1.6 Reading test and questionnaire data analysis	28
3.2 Lexile analysis of reading materials	28
3.3 Interviews with lecturers	29
3.3.1 Interview participant recruitment	29
3.3.2 Final interview participant sample	29
3.3.3 Interview guidelines and procedure	30
3.3.4 Interview data analysis	31
4. RESULTS	32
4.1 Research Question 1	32
4.1.1 Research Question 1a	37
4.2 Research Question 2	38
4.2.1 To what extent do students perceive the texts to be difficult?	38
4.2.2 How do lecturers ask students to work with these texts?	43
4.2.3 What kind of support do students get?	47
4.2.4 How can course instructors explain gaps between reading demands and reading abilities?	51
5. DISCUSSION	53
5.1 How do students' proficiency and measures of text complexity compare?	54
5.2 What makes texts difficult?	55
5.3 Which reading tasks are EMI students asked to do?	56
5.4 What kind of support do students get?	57
6. LIMITATIONS OF THE CURRENT STUDY	58
7. AREAS FOR FUTURE RESEARCH	59

8. CONCLUSION	60
REFERENCES	61
Appendix A: Student questionnaire	67
Appendix B: Test administration guidelines	72
Appendix C: Information document for reading test participants	74
Appendix D: Information document for interview participants	75
Appendix E: Consent form for study participants	76
Appendix F: Biodata questionnaire and interview guidelines	77

LIST OF TABLES

Table 1: Minimum score (mean) necessary to be admitted to EMI programmes (Ferencz et al., 2014, p. 73)	10
Table 2: Selected Lexile text measure ranges in secondary and post-secondary education	16
Table 3: Number of institutions per country with EMI bachelor's programs that were contacted.....	21
Table 4: Countries of institutions, number of EMI programmes, and students participating	23
Table 5: Participants' gender	24
Table 6: Participants' age.....	24
Table 7: Participants' languages	25
Table 8: Participants' year of study	26
Table 9: Participants' intended major.....	26
Table 10: Reading materials included in the analysis for each institution.....	28
Table 11: Lecturers' age, experience, and subject areas	30
Table 12: Participants' Aptis Reading Test scores and corresponding CEFR levels and Lexile reading measures	33
Table 13: Lexile text measures for each institution	34
Table 14: How similar did students think the texts in the Aptis reading tests were to the texts they need to understand in their EMI university courses?	37
Table 15: Why do students find the reading materials difficult?	39
Table 16: How do students cope with difficult reading materials in their courses?	47

LIST OF FIGURES

Figure 1: Boxplot comparison of Lexile measures between students and texts	35
Figure 2: Comparison of Lexile measures between students and texts for the institution in Egypt ..	36
Figure 3: Comparison of Lexile measures between students and texts for the institution in Lithuania.....	36
Figure 4: Comparison of Lexile measures between students and texts for the institution in Austria	36
Figure 5: How difficult do students find the reading materials in their courses?	38

1. BACKGROUND AND LITERATURE REVIEW

1.1 The role of English as a medium of instruction (EMI) in international academia

In academia, EMI can be defined as “the use of the English language to teach academic subjects in countries or jurisdictions where the first language (L1) of the majority of the population is not English” (Dearden, 2014). Teaching academic subjects through the use of the English language has seen a steady increase in popularity over the last decades (Dearden, 2014; Ferencz et al., 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 at postgraduate and, to a lesser extent, undergraduate levels. In 2016, in Austria alone, the number of programmes offered were 119 EMI master’s, 44 EMI PhD and 3 EMI undergraduate (i.e., bachelor’s) programmes (Austrian Federal Ministry of Education, Science and Research, 2018). Wächter and Maiworm (2014), who investigated the spread of EMI programmes throughout Europe, found that the number of programmes increased from 725 in 2001 to 8,089 in 2014. They observed highest absolute numbers in the Netherlands, Germany, and Sweden, but also a strong increase in Central Eastern Europe and the Baltic states. However, only 1.3% of all students in Europe were enrolled in EMI programmes in the academic year of 2013/14, with 54% of these being foreign students, which underlines the importance of EMI courses for internationalisation. Programmes provided exclusively in English generally tend to be more popular with universities of applied sciences, which are often privately funded and comparatively small in size (Frank et al., 2010; Gürtler & Kronewald, 2015). This might explain the comparatively low proportion of students enrolled in EMI programmes.

In China, the rise of EMI began in 2001 when government policies included the number of EMI courses as a performance indicator for the evaluation of universities (Galloway et al., 2017). By 2006, 132 of 135 universities across China were offering EMI courses or programmes, as Wu et al. (2010) revealed in a national study (as cited by Hu et al., 2014, p. 22).

The growth of EMI programmes since 2000 has happened at an even faster speed in Japan (Macaro et al., 2018). Similar to the development in China, this was the consequence of state policies that intended to increase the internationalisation and competitiveness of Japanese higher education. Macaro et al. (2018) point out that the rise of EMI is not universally welcomed in all Asian countries. They mention Bangladesh as an example, where English (the language of British colonial rule) is used in the private sector of higher education but not allowed in public universities.

Although Coleman (2006) states that in Africa, English was already established as the language of higher education in the early 2000s, there is little research into EMI in this part of the world. Our literature search revealed only one study, which found that EMI is widespread in South African and Namibian schools (Uys et al., 2007). However, we were not able to find published empirical research on the use of EMI in other regions of Africa, nor on the relationship between text complexity and student proficiency in EMI higher education in Africa. This geographical area thus seems to be fruitful ground for novel insights into this area.

1.1.1 Reasons for offering EMI courses and associated challenges

The reasons for offering EMI courses are manifold. Most often, they are rooted in a perceived need for internationalisation (Clifford & Montgomery, 2011), which includes strengthening an institution's international profile (Maiworm & Wächter, 2014), increasing staff mobility and attracting more international (i.e., higher paying) students (Coleman, 2006), as well as highly qualified international university staff and PhD students (Maiworm & Wächter, 2014). In turn, EMI courses are also considered to increase the chances for local students in international academia and business and increasing the level of English within the home country. Another argument is the prestige of the English language, resulting in access to better employment for individuals and in access to prosperity and modernity on a larger scale (Dearden, 2014).

The increasing role of English as a lingua franca in academia can be considered a driving force in the rise of EMI courses worldwide, but there are challenges with such programmes. Coleman (2006) addresses the potentially inadequate level of English proficiency among domestic and international students. However, according to a survey by Lam and Maiworm (2014), EMI programme directors state that the main problem for teaching in EMI courses is the students' heterogeneity in terms of English language proficiency. Lam and Maiworm further note that although EMI instructors generally have a good command of English, this does not necessarily mean they can handle and adapt to the wide range of language levels in their classroom. Surprisingly, despite the challenges associated with heterogeneity in terms of language skills, the study also found that one third of all EMI programmes and institutions included no language training in their programme. This observation appears to be slightly more pronounced in postgraduate, rather than undergraduate, programmes.

EMI programmes have spread in diverse contexts, and for many countries, especially those with a colonial past, questions of language policy may be ideologically charged. An example of this is the aforementioned Bangladesh. The use of English as a lingua franca for higher education is sometimes criticised as linguistic imperialism and Westernisation (Galloway et al., 2017).

The limited availability of level-appropriate teaching materials in English across different subject areas presents an additional challenge (Coleman, 2006) and consequently, British or US American textbooks, with their respective cultural settings, are often used. Dearden (2014) suggests that some countries may be hesitant to introduce EMI programmes because they may have a negative effect on the national language. Some stakeholders consider the learning of English as a threat to national, cultural and linguistic identity (Macaro et al., 2018; McIlwraith & Fortune, 2016). Despite these critical voices, EMI courses have been steadily on the rise over the past decades and this will likely continue over the coming years.

1.1.2 Subject areas and admission criteria for EMI courses

The results of a survey carried out in Germany among university teaching staff showed that EMI programmes were offered across a range of subject areas, with business and economics programmes being the frontrunner in terms of percentage of EMI courses, and engineering programmes in terms of absolute numbers (Gürtler & Kronewald, 2015). Brenn-White and Faethe (2013) also found a similar distribution of disciplines among master's programmes over Europe. In their study, the third most popular discipline was the field of social sciences. Brenn-White and Faethe report that the representation of subjects among EMI programmes has remained stable over the last few years. Wächter and Maiworm (2017), however, report the highest proportion of EMI programmes in social sciences, business and law (35%), followed by sciences (23%) and engineering, manufacturing and construction (18%).

Among the criteria used for admission to EMI programmes at bachelor's level, English proficiency was applied as a selection criterion for college readiness in the majority of cases. Roughly 90% of EMI programmes ask for proof of English proficiency in form of international certification. Language proficiency is thus one of the most important admission criteria to EMI programmes alongside the academic, intellectual and/or artistic skills required for the respective degree programme (Ferencz et al., 2014). Ferencz et al. (2014) compared the mean minimum scores for admission across 427 undergraduate and postgraduate degree programmes that used large-scale international tests as entrance requirement. As shown in Table 1, for two of the most commonly-used tests for EMI course admission, the Test of English as a Foreign Language (TOEFL) and the International English Language Testing System (IELTS), the majority of candidates need to achieve scores equivalent to B2 on the Common European Framework of Reference for Languages (CEFR) to be eligible to enrol in a bachelor's or master's programme. Only a small number of bachelor's programmes accept scores equivalent to CEFR B1; however, even for these programmes, the entry requirement is very close to the cut-off to B2, which is set at 542 for the TOEFL ITP (Educational Testing Service, 2020). Thus, overall, the entrance criteria in Table 1 suggest that that successful students will be able to cope with the demands of university material (Carlsen, 2018).

Table 1: Minimum score (mean) necessary to be admitted to EMI programmes
(Ferencz et al., 2014, p. 73)

Examination	Bachelor	CEFR equivalent	Master	CEFR equivalent
TOEFL iBT	79	B2	86	B2
TOEFL ITP	538	B1	563	B2
IELTS	6	B2	6	B2
Number of programmes	66		361	

CEFR equivalents based on Papageorgiou et al. (2015) for TOEFL iBT, Educational Testing Service (Educational Testing Service, 2020) for TOEFL IPT, and IELTS (IELTS, 2020) for IELTS

The Aptis test, which is not intended for use in admissions, is not included in Ferencz et al.'s (2014) table above. However, Aptis test scores have been aligned to the CEFR in a four-stage CEFR linking project, as described in O'Sullivan (2015). Scores in the four skills are reported to test-takers based on a scale ranging from 0–50 points and a maximum of 200 points. In addition, individual CEFR levels for each component are calculated and then combined to form an overall CEFR level. Dunn's (2020) technical report provides a helpful overview of how numerical cut-off scores for individual skills have been linked to the CEFR. The cut-off score for reading in the Aptis test is set at 38 out of 50 points for B2 and 44 points for C1 or above.

1.2 The role of reading in EMI university settings

Reading is perhaps the most fundamental academic skill, and alongside writing, critical thinking, oral presentation and media literacy skills, it is a prerequisite for academic success in higher education (Barth, 2003; Cox et al., 2003; Hermida, 2009). It has been shown that 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). Students who struggle with reading might be unable to take more challenging courses (Au, 2000) and may consequently be hindered from acquiring a degree (Schmeiser, 2009). It is thus not without reason that university students are said to “read for a degree” (Aldridge, 2019).

When reading critically for academic purposes, students need to display a number of subskills. Although the necessary subskills depend on the purpose of the reading activity, students in university settings are generally required to be able to: (a) identify patterns of textual elements; (b) differentiate between main and secondary ideas; (c) evaluate the reliability of the source; (d) analyze the argumentation line; and (d) make inferences about the reading materials (Manarin et al., 2015, p. 47). However, explicit teaching of academic reading skills and enhancing academic literacy, although central to academic success, can be rarely found in university curricula (Hermida, 2009).

Well-developed reading skills are likewise crucial in contexts of L2 usage in reading for academic purposes within EMI instruction (Owen et al., 2021). L2 students of English, however, typically face additional challenges in reading for academic purposes 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 (Mezek, 2013). Furthermore, the nature of reading strategies and purposes seem to have an impact on text difficulty. For example, expeditious reading has been shown to be more challenging for L2 readers than careful reading (Weir et al., 2000).

Despite the importance of reading in university settings (including EMI contexts), as well as the rising number of EMI programmes and the above-discussed challenges of reading in an L2, there is to date, only scarce knowledge about reading proficiency and academic reading skills in EMI universities.

1.3 Testing reading skills as part of EMI university admission tests

To enrol for EMI courses at a university, students generally need to pass a language test. Such university entrance exams need to meet a number of 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 terms of both situational authenticity (the degree to which the test tasks reflect the context of target language use [TLU] tasks), as well as interactional authenticity (the degree to which the processes needed to succeed on the test tasks reflect the processes needed for TLU tasks). Put differently, university entrance tests and score interpretations which are based on them should be valid in that they are *generalizable* to the target-language use domain outside the test situation. In the case of university entrance tests, authentic language tests should reflect the critical features of relevant real-life academic activities by choosing content and tasks that future students will be likely to encounter in a non-test academic situation. As reading is one of the core activities in higher education, university entrance tests should mirror, as much and as closely as possible, the reading activities that students will encounter during their future education or training.

A number of large-scale tests and test scores are currently used by EMI universities worldwide to decide whether to admit students. Owen et al. (2021), for instance, analyzed the suitability of the TOEFL iBT reading section as an entry test to two different EMI programmes in Sweden and Nepal, thereby targeting two different socio-economic and geographical contexts. The authors identified three purposes of academic reading that the TOEFL iBT reading test aims at: (a) reading for basic comprehension; (b) inferencing; and (c) reading to learn. Their study demonstrated that, overall, the TOEFL iBT reading test appeared to be a suitable instrument for measuring academic reading skills as a pre- or post-entry language test in both EMI contexts.

Nonetheless, Owen et al. (2021) caution that different EMI contexts may require localisation in terms of different academic reading skills and vocabulary knowledge. Other tests, such as the Aptis General (O'Sullivan et al., 2020) and the Aptis Advanced (British Council, 2020c), also purport to assess both careful and expeditious reading competences in TLU situations “where learners are engaged with real-world tasks in higher education and training programmes” (British Council, 2020c, p. 4).

Yet, authenticity, and thus validity, applies not only to reading purposes in an entrance test but also to complexity of reading materials. In other words, the complexity of reading materials in a test suite should resemble those in real-life academic texts. Until the turn of the last century, claims about the complexity of reading materials used in such examinations had to rely on expert judgements, due to the lack of reliable quantitative tools to investigate textual features of a larger body of material (Biber et al., 2004). College transcripts, as well as course completion and dropout rates, have been used as traditional proxies for university readiness (McCormick, 1999), rather than selecting students based on valid and authentic entrance examinations. The last decade has seen drastic advances in computational linguistics, which have enabled more 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), for instance, have investigated the relationship between reading passages in the IELTS examination and textbooks encountered by students in their first year at university. They employed a range of mechanical measures, as well as expert judgements, to gauge potential gaps between what students can understand and what they would be expected to understand at university. University reading materials seemed to be marginally more challenging than IELTS reading passages at surface level, as indicated by 24 out of 29 measures. For instance, the IELTS texts contained a slightly higher percentage of frequent words (76.84% from the first 1K from the BNC as compared to 74%) and on average fewer academic words (7.9% of the words found in IELTS material also appeared on the AWL, compared to 10.51%), as well as fewer words not found within the first 15K of the BNC (1.09% versus 4.33%). The latter results, as well as the analysis of expert judgements, indicated that the major difference between university textbooks and IELTS materials is the degree of cultural and subject specificity. Overall, however, the authors conclude there are “few fundamental differences between the texts that students might expect to encounter in their first year of study and those used in the IELTS academic reading test” (Green et al., 2010, p. 207). Thus, studies such as this one are an important step in establishing a validity argument for language tests used in university admission.

1.4 Measuring text complexity using the Lexile Framework® for Reading

From the 1920s onwards, there has been an ongoing interest in measuring the complexity of a text and determining its *readability* for a specific audience (DuBay, 2004; Stenner et al., 2007). Dale and Chall (1949) offer a comprehensive definition of readability as “the sum total (including the interactions) of all those elements within a given piece of printed material that affects the success that a group of readers have with it”, meaning that the readers “understand it, read it at an optimum speed, and find it interesting” (p. 23). Over the last century, a considerable number of studies have researched readability, resulting in mathematical readability formulas being suggested and continually refined (DuBay, 2004; Stenner et al., 2007). These statistically-based equations have then been applied on a large scale – sometimes even beyond their original purposes – to objectively measure the comprehension difficulty of texts and to develop writing guides in education, research, journalism, and public services (Crossley et al., 2017; DuBay, 2004).

Research largely agrees that sentence complexity and the familiarity of vocabulary are relevant factors to predict text complexity. Still, their operationalization in readability formulas varies. The different readability formulas represent different attempts to measure text difficulty in mathematical operations, using a variety of lexical (e.g., number of affixes, syllables, off-list words, pronouns, monosyllabic words, or words counting three or more syllables) and syntactical (e.g., words per sentence, number of sentences per 100 words) variables in their calculations. These variables are deemed proxies for either sentence complexity and lexical difficulty (see DuBay, 2004, for a review of relevant literature). Amongst the most frequently used readability formulas in English are Flesch’s Reading Ease Formula (1948), the New Dale-and-Chall formula (Chall & Dale, 1995), the Fry readability graph (e.g., Fry, 1968), the Fog Index (Gunning, 1952), or the Flesch-Kincaid Grade Level, which is also in-built in *Microsoft Word* software (Kincaid et al., 1975). For the purposes of this study, the Lexile Framework for Reading®, which is “the most widely used reading metric in the world” (MetaMetrics, 2020a; Swartz et al., 2011) will be of particular interest.

On the market for over 30 years, the Lexile Framework for Reading®, developed by MetaMetrics Inc., is a widely-used readability index for measuring reading ability and text complexity. Students from 180 countries receive Lexile measures and in the US alone, Lexile reading measures are reported annually to over 35 million students across all states. Additionally, more than 100 million online and offline reading materials now indicate their Lexile text measure for educators, parents, test developers, and readers themselves (MetaMetrics, 2020a). The Lexile scale is a developmental interval scale based on the Rasch model and grounded in readability theory, school practice, and educational science (MetaMetrics, 2020a; Stenner et al., 2007; Wright & Stenner, 1999). Similar to the majority of readability measures, the Lexile Framework for Reading includes semantic (i.e., word frequency) and syntactic (i.e., sentence length) factors to measure text complexity (Mesmer, 2008; Stenner et al., 2007).

Unlike other readability formulas, the Lexile Framework uses word frequency as a proxy for similar variables, such as numbers of letters or syllables of words (Stenner et al., 2007; Stenner & Burdick, 1997). Originally based on the *American Heritage Intermediate Corpus* (Carroll et al., 1971) with 5 million words from 1,045 pieces of reading materials (Stenner & Burdick, 1997), the average word frequency count has been extended to a 600-million word entry corpus (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). Thus, MetaMetrics strongly promotes its advantage of combining reading ability in English and text complexity, i.e., reportedly two of three important factors of text comprehension alongside the purpose of reading (Stenner et al., 2007). Lexile reading measures 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 programs. For illustration, the low-end of the scale is situated at the average difficulty of primer reading materials, whereas the upper-end of the scale is defined by the mean difficulty of technical encyclopaedia entries (MetaMetrics, 2020c; Stenner et al., 2007; Stenner & Stone, 2004). To achieve an optimum match between reader and text, MetaMetrics advises teachers, parents and students to choose reading materials that are no more than 50L above or 100L below the reader's Lexile score (MetaMetrics, 2020c).

A Lexile measure match between a reader and a text forecasts a 75 % *target comprehension rate*. Put differently, if readers with a reading ability of 600L are assigned reading materials calibrated equally at 600L, the framework predicts them to understand approximately 75 % of the materials. If, though, the text's Lexile text measure is too high and the comprehension rate is situated at 50 % only, the amount of unfamiliar vocabulary and sentence structure will result in short-term memory overload and thus in insufficient comprehension and learner demotivation (Stenner & Burdick, 1997). The cut-off score at 75% comprehension rate is purportedly "highly useful" (Stenner & Stone, 2004, p. 14), although it may appear somewhat arbitrary. Therefore, the Lexile framework provides an "interpretative – [and] not prescriptive – tool for measuring student growth and predicting future success" (MetaMetrics, 2020b).

Nevertheless, readability literature highlights certain restrictions of readability formulas in predicting and measuring text comprehension in general and the Lexile Framework in particular. First, the Lexile formula is limited to continuous prose, and it has been demonstrated to be insufficient in predicting the complexity of either poetry or non-continuous prose texts (e.g., recipes, shopping lists) (Stenner & Burdick, 1997). DuBay (2004) further criticizes the Lexile Framework for excluding contextual factors, such as background knowledge and subject-specific knowledge, which may enhance or hinder text comprehension.

Most importantly, however, the fact that readability measures such as the Lexile Framework are based on L1 reading abilities implies that these may not consider or be applicable to L2 reading skills (L2). While Carrell (1987) called for "a clearer theoretical approach to readability, one that takes a broader range of reader as well as context variables into consideration" (p. 34) (in particular whether reading in an L1 or an L2), her call appears to have remained largely unanswered over the last 30 years.

In the same vein, Beinborn et al. (2014) state more recently that L1 acquisition and L2 learning processes are unlike, and L2 learning frequently follows L1 acquisition, with L2 readers relying on linguistic resources at their disposal and previous reading experience. In other words, L2 readers tend to be older and thus cognitively more advanced than (mainly younger) L1 readers. Yet, L2 readers are usually disadvantaged with regards to vocabulary input. In sum, applying the same readability measures to both L1 and L2 without reflection nor adaptation is problematic, according to Beinborn et al. (2014). Thus, further research in second language readability and readability formulas is clearly necessary.

1.5 Previous research on university reading materials using the Lexile Framework®

In recent years, a number of 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 addressed the “gap in literacy requirements”, or *reading gap*, between high school and post-secondary reading materials in the American educational context (see Table 2). The study compared reading materials used in the last two years of secondary education in the US with texts from core subjects in humanities and social science courses in the first two years of university or college in the US. The author also looked at occupational reading materials from 16 career clusters, as well as materials encountered in the military and public domain. Williamson reports a monotonic increase in difficulty from secondary education to military texts, and further to workplace and university reading material. Texts from the first two years of university study centred around a median of 1355L. In a later study, Williamson and colleagues (2016) found similar Lexile levels for reading materials used in universities across the UK, however the authors do not specify which years of university study they investigated (i.e., texts for beginner students, or texts for all students combined, but the study design indicates that it is likely the latter).

Other studies identified lower Lexile levels for post-secondary reading. For example, Lexile text measures reported by Wilkins et al. (2010) for US university reading materials are similar to those found by Williamson et al. (2012) for materials used in grade 12 in the US. These studies reported text measures of 1144L and 1130L, respectively. Importantly, however, Wilkins et al. (2010) based their analysis specifically on textbooks used by beginning university students of English, rather than on a combination of texts for both first and second year students (as in the study by Williamson, 2004). Wilkin’s et al’s context is thus comparable to the context of 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 UK, with a median of 1030L. Table 2 presents an overview of these studies.

Table 2: Selected Lexile text measure ranges in secondary and post-secondary education as cited in earlier studies

	25 th -percentile	Median	75 th -percentile	Study
Universities (year 1 and 2 combined) (US)	1253L	1355L	1450L	Williamson (2004) N=150 text collections
Universities (year unspecified) (UK)	1210L	1310L	1360L	Williamson et al. (2016) N=99 text books
Universities (year 1 beginner) (US)	1106L	1144L	1265L	Wilkins et al. (2010) N=74 text books
Grade 12 (US)	1110L	1130L	1220L	Williamson et al. (2012) n=34 text books
Key Stage 4 (UK) (years 10 and 11)	920L	1030L	1110L	Koons et al. (2016) N=29 text books

All of the above studies utilising the Lexile Framework have been conducted in English-speaking countries; however, considering the increase in EMI programs 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. Although a lack of linguistic abilities to cope with the demands of EMI programmes is seen as a major concern by stakeholders (Lam & Maiworm, 2014), there seems to be a lack of empirical studies on this issue. This is particularly concerning as Lexile measures are normed on L1 speakers of English. Only a small number of studies have investigated college readiness in terms of readability measures in EMI contexts outside the English-speaking world. 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 post-secondary 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-speaking countries, and the important role of reading within these programmes, the present study aims to address this by investigating the potential gap between reading demands of course texts and reading abilities of EMI students, and the ways in which course tutors perceive and mitigate such gaps.

2. RESEARCH QUESTIONS

It emerged from the review of previous research that there are two major gaps in the literature on reading demands and text complexity. First, even though most EMI institutions demand proof of language proficiency, and reading skills are crucial for academic success, there is scant knowledge about students' L2 reading skills in EMI programmes. Second, studies into text complexity of university reading materials using Lexile text measures have hitherto been mostly confined to L1 educational contexts of English, paying little attention to the global and ever-growing trend of EMI.

To address this, the main aim of the present study is 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 reading measures). Students' reading ability was assessed using two different versions of the Aptis Reading Test (for a description of the two tests and the rationale behind using two different versions, see Section 3.1.1). The advantage of using the Aptis tests was that its scores have been formally linked to the Lexile Framework for Reading (the linking study was made available to us by Aptis, but unfortunately it cannot be discussed here for reasons of confidentiality). 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. Thus, the Aptis test allowed us to establish participants' Lexile score, based on which we were then able to compare EMI students' reading skills to in-use university reading materials based on the Lexile scale.

The following research questions are addressed:

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

RQ1a: How similar do students think the texts in the Aptis reading tests are to the texts they need to understand in their EMI university courses?

RQ2. If there is a mismatch between EMI students' scores and the textual demands in EMI courses as estimated by Lexile text measures,

- a) To what extent do students perceive the texts to be difficult?**
- b) How do lecturers ask students to work with these texts?**
- c) What kind of support do students get?**
- d) How can course instructors explain this gap?**

We investigated the research questions through a mixed-methods design, which is outlined in the following section.

3. METHODOLOGY

The research questions were answered by three main data sources. The first data source were Aptis reading test results with corresponding Lexile reading measures, along with questionnaire responses of EMI university students across three different universities. The second type of data were Lexile text measures of EMI university reading materials which the participating students were required to read as part of their studies. And finally, the third data source were interviews with lecturers who used the reading materials with the participating students in class. Collecting these different types of data allowed us to cross-compare and triangulate the findings. The data collection procedures for each data source are described below.

3.1 Reading test and questionnaire

For this part of the investigation, students from three EMI university programmes completed the Aptis reading test and filled in a questionnaire in order to inform RQ1, RQ1a, and RQ2. The reading test and questionnaire will be outlined in more detail below, before describing the different phases of participant recruitment and the final group of participants. This section then outlines the test administration procedure and concludes with a brief description of the data analysis.

3.1.1 Aptis reading tests

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, 2020a). The tests include a listening, reading, speaking, writing, and grammar/vocabulary component. In the current study, only the reading components of the two tests were used. It should be noted that we used older versions of the tests which have since been retired in the case of Aptis General, as the British Council launched revised test versions in April 2020 (which 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 our study may have been slightly different if the newer versions were used.

The Aptis General Reading Test used in the study targeted CEFR levels A1 to C1 and consisted of four different parts. In Part 1 candidates had to complete gaps in five sentences by choosing the correct word for each gap from a pull-down menu. In Part 2 they had to put eight sentences from a linear text in the correct order, whereby only the first sentence was given. Part 3 contained a longer text (about 150 words) and candidates again had to fill gaps by choosing the correct word for each gap from a list. Finally, in Part 4 candidates had to read a longer text consisting of seven paragraphs and then choose the correct heading for each paragraph from a list of eight options.

The Aptis Advanced Reading Test used in the study “ha[d] the optimal point of discrimination at levels C1 and C2” (British Council, 2020b) and also consisted of four different parts. In Part 1 candidates first read four texts about different people and then had to match seven statements to the individual people. Part 2 was similar to Part 4 of the Aptis General Test described above, in that candidates had to match a total of eight headings with seven paragraphs within a longer text, whereby one of the eight headings served as a distractor. In Part 3 candidates read a longer text of about 300 words containing five gaps and they had to choose the correct words or phrases from a pull-down menu for each gap. Similarly, in Part 4 candidates read two shorter texts with a total of six gaps and they were asked to identify the correct words and phrases for each gap from a pull-down menu.

Using both tests instead of just one had two main advantages. First, both reading tests have been formally linked to the Lexile Framework for Reading as outlined above, so students taking the tests receive Lexile reading measures in addition to their test results. Comparing the Lexile reading measures generated through the tests with the Lexile text measures of the university reading materials allowed us to draw conclusions with regards to which of the two tests may be more suitable for university admissions purposes. And second, as the reading tests have also been linked to the CEFR, and students are placed on a CEFR level depending on the number of points they score, we could compare the CEFR scores of students taking the Aptis General Test with the CEFR scores of students taking the Aptis Advanced Test, which again may serve as an indicator as to which of the two tests is better suited for university admission purposes.

3.1.2 Questionnaire

After completing the Aptis test, students filled in an online questionnaire (see Appendix A). The questionnaire was designed by the researchers based on the research aims. The first version of the questionnaire was piloted on five EMI university students, and piloting indicated that the wording of two questions required more specificity. Questions relating to the degree program were changed to refer to the candidates' intended major, as not all participants would have entered a particular degree program yet and would still be in the more general, foundational courses that their institution assigned upon admission. Also, questions on candidates' L1 and their year of study were added to the questionnaire to obtain more detailed information on the participants' profiles.

Before answering the individual items, students first had to enter their candidate reference number (as shown on their keycode slips, see Section 3.1.5 below), which allowed us to link their questionnaire responses with their test results. This was important for the analysis of the data. Following this, students answered three items on biodata (gender, age, and first language) and two items on their intended major (which major they were studying and which year they were in).

The questionnaire then included items on the university reading materials that students typically encountered during their studies, in order to inform RQ2. In the first of these items, students had to indicate on a four-point Likert scale how difficult they found the reading materials in the courses for their intended major (very easy, rather easy, rather difficult, or very difficult).

Depending on which response they chose, the next item then asked them why they found the reading materials easy or difficult. For this question, students could choose from a number of options and they could also note down additional reasons as an open answer. Students who indicated that they found the reading materials for their courses rather difficult or very difficult were also asked how they coped with difficult reading materials, whereby they could again choose from a number of different options (e.g., ask the course instructor, ask student colleagues, use a dictionary, etc.) and also note down comments to add responses not included in the pre-defined list of answers. In the final item in this section, students who found the reading materials difficult could note in what ways their course instructors helped them with the texts.

In order to inform RQ1a, the following question asked all students how similar they found the texts in the Aptis reading test compared to the texts in the courses for their intended major. Students had to indicate their perceived level of similarity on a five-point Likert scale, where 1 was defined as “very similar”, 5 was defined as “very different”, and the intermediary categories (2, 3, and 4) remained undefined.

In the final item students were asked to self-assess their English reading proficiency by choosing which descriptors about their ability to read texts in English applied to them. The four descriptors were taken verbatim from levels B1 to C2 (one descriptor consisting of one or two sentences for each level) of the CEFR self-assessment grid (Council of Europe, 2001). Students were asked to tick all descriptors they thought applied to them.

3.1.3 Reading test participant recruitment

The first step in collecting data was to establish contact with potential partner institutions to help with student recruitment. This happened in two phases. During the first phase (September 2017 to October 2018) EMI programmes at bachelor’s level across Europe were identified and contacted. However, it proved very difficult to recruit students from these programmes at the beginning of their studies, either due to the size or novelty of the programmes themselves, or because of temporal constraints or lack of human resources at the respective institutions to receive internal ethical clearance, compile texts, and administer the reading test. A small number of programmes reported that they were already in the process of in-house evaluation and therefore had to decline. As a result, the vast majority of the institutions we contacted did not participate in the study, so the target population had to be widened to also include master’s degree students. In addition, we decided to target a number of universities outside of Europe through contacts provided by the British Council. During the second phase (November 2018 to December 2019), bachelor’s and master’s level EMI programmes were contacted and a number of institutions participated in the study. The two phases of recruitment are outlined in more detail in the following section.

3.1.3.1 First phase of recruitment

Initially, potential partner institutions were expected to send textbooks and reading materials via post or scan them in situ and send them online, and provide staffing for, as well as administer the Aptis reading test on site. As sending or scanning material was identified as a potential barrier for institutions to participate, we asked them to compile material in the form of a list of texts that would then be ordered or scanned at our institution.

As a first step to acquire partners, suitable bachelor's level EMI programmes across Europe were identified (i.e., programmes that were fully taught in English). Thus, the study excluded programmes taught predominantly, but only partly in English. Programmes were identified mainly through the database studyportals.com, which provides information on 79,793 bachelor's programmes worldwide and has been used in earlier studies and surveys to identify specific programmes (Maiworm & Wächter, 2014). Table 3 shows the number of all institutions with bachelor's programmes in English contacted by email and phone. The information available online was used to contact the heads of the departments or the programme coordinators, wherever possible. In a small number of instances, a generic contact address had to be used as the contact details of the programme coordinators was not available online.

Table 3: Number of institutions per country with EMI bachelor's programs that were contacted

Country	N
Austria	11
Belgium	9
Denmark	5
France	14
Germany	23
Lithuania	1
Netherlands	13
Spain	20
Sweden	38
Switzerland	6
Turkey	3
Total	143

Of the 143 institutions and study programmes, the majority did not react or declined immediately, while a handful of institutions expressed interest in the study. After discussing the study within the faculty, however, most of these decided to drop out for the reasons mentioned above (mostly lack of resources, temporal constraints, ongoing in-house evaluation, or difficulty in receiving internal ethical clearance). One university in Lithuania agreed to take part and set up test dates with approximately 100 students. They also compiled and sent reading materials to us electronically. However, the contacts at the university eventually decided to retract participation due to a malfunction of the testing platform on the day of the first administration.

3.1.3.2 Second phase of recruitment

After widening the target population to also include master's programmes, we first contacted institutions in and around Innsbruck to take part in the study. We felt that participation would be more attractive for institutions if we could administer the test and collect the materials ourselves, rather than ask the institutions to do this. One institution in Innsbruck offered seven different EMI master's programmes, and we contacted the director of study for each of these programmes to ask whether they would be interested in taking part. Contact was established with three directors of study, and they were all interested; however, they left it to the students to participate on a voluntary basis. In order to attract students to participate, we presented the study to one group of 28 students at this institution with the incentive that they would receive their test scores (including corresponding CEFR levels) some weeks after taking part. However, at the day of the administration only 5 out of the 28 students took the test, which unfortunately led the dean of the institution to cancel all subsequent administrations for the remaining approximately 120 students, arguing that students would be too busy as the end of semester was approaching. For this reason, the institution also declined to provide information on the reading materials, so the 5 students that were tested could not be included in the final sample.

A second programme at the University of Innsbruck offering one EMI master's course with 20 students also agreed to participate after establishing contact through the dean of studies. This institution again had to leave it to the students to take part on a voluntary basis, as the research project did not match the course curriculum closely enough to warrant compulsory participation. Due to the low turn-around of the prior administration, we decided to offer students a 10€ inconvenience allowance as well as their personalized Aptis test results to potentially increase the number of participants. For this administration, 15 students took the test. Out of these, 13 students could be included in the final sample for analysis because 2 students did not fill in the online questionnaire (see also the test administration procedure outlined below).

As the monetary incentive seemed to make voluntary participation more attractive for students, we re-established contact with the university in Lithuania from the first phase of recruitment, as we had already received the scanned reading materials for the EMI bachelor's courses offered at this university. Despite their unsatisfactory experience during the first administration due to the malfunction of the testing platform on the day of the test, they agreed to participate a second time, with a potential number of 120 students across several EMI bachelor's programmes. After discussing the monetary incentive with the university contact, they suggested 5€ for participating students, as 10€ was perceived to be an unusually high sum for this kind of research study in Lithuania. To further maximise turnout, 14 different test slots at different times of day were offered in October 2019, and students could sign up for any of the 14 slots. Participating students were also entered in a raffle with the chance to win university merchandise materials such as tote bags and water bottles. Despite these efforts, only 30 students across four different EMI bachelor's courses took part in the study. Out of these, 26 could be included in the final sample for analysis. The remaining 4 students either did not fill in the online questionnaire or entered an incorrect candidate reference number in the questionnaire, which made it impossible to link their test performance with their questionnaire responses.

In light of the limited insights into EMI university education in Africa (see discussion in Section 1.1) and the need for further participants due to negotiations with other potential partner universities in Europe proving challenging, we decided to try and include an African institution in the study. As the British Council had collaborated with a university in Egypt offering several EMI bachelor's courses in the past, a BC representative facilitated contact with this university. With the help of two lecturers at the university, we organized a number of test administrations with EMI students. After receiving university internal approval for the project, students were recruited through online student portals and through their course instructors. Students were again offered 5€ for taking part. From this university, a total of 84 students across 15 different EMI bachelor's programmes participated in the study, out of which 68 could be included in the final sample. Of the remaining 16 students, 10 students only completed the first of four reading tasks (possibly due to problems with internet connectivity), 2 students did not submit the test responses so their test results could not be generated, and 4 students provided incorrect candidate reference numbers in the questionnaire so their test results could not be linked to the questionnaire responses.

Thus, the final sample included in the analysis consisted of 107 students across three institutions and 18 different EMI bachelor's and master's programmes in Europe and Africa, as shown in Table 4. We initially planned to test a more homogenous sample and also a larger number of students, however, recruiting participants proved very difficult for the reasons outlined above.

Table 4: Countries of institutions, number of EMI programmes, and number of students participating in the study

Country of institution	No. of EMI programmes	Bachelor/Master	No. of students
Egypt	15	Bachelor	68
Lithuania	4	Bachelor	26
Austria	1	Master	13

3.1.4 Final reading test participant sample

As outlined above, the participants were 107 students from three universities offering EMI programmes in Egypt ($N = 68$; 63.6%), Lithuania ($N = 26$; 24.3%), and Austria ($N = 13$; 12.1%), as shown in Table 5. Across all institutions, almost two-thirds of the students were female ($N = 67$; 61.6%) and slightly more than a third were male ($N = 39$; 37.9%); one student identified as non-binary. The Egyptian and Lithuanian samples included a comparatively larger proportion of female test-takers (61.8% and 69.2% respectively), whereas the Austrian participants were more evenly distributed in terms of gender (female = 53.8%, male = 46.2%).

Table 5: Participants' gender

Country of institution	Gender	n	%
Egypt	female	42	61.8
	male	25	36.8
	non-binary	1	1.5
Lithuania	female	18	69.2
	male	8	30.8
Austria	female	7	53.8
	male	6	46.2

Across all institutions, the majority of test-takers were younger than 21 years (see also Table 6; note that students could choose between “15–20”, “21–25”, “26–30”, “31–35”, and “older than 35”; see also Appendix A). Egyptian and Lithuanian bachelor's students were predominantly between 15 and 20 years old, with only two Egyptian students aged 21 to 25 years. Unsurprisingly, the Austrian master's students were generally older, with their ages ranging mainly between 21 to 25 years (61.5%) and 26 to 30 years (30.8%). One Austrian student was in their early thirties.

Table 6: Participants' age

Country of institution	Age	N	%
Egypt	15-20	66	97.1
	21-25	2	2.9
Lithuania	15-20	26	100
Austria	21-25	8	61.5
	26-30	4	30.8
	31-35	1	7.7

As outlined in Table 7, the sample included a considerable number of multilingual speakers. All test-takers in the Egyptian sample reported to be L1 speakers of Arabic ($N = 68$) and a substantial sub-sample indicated that they were speakers of additional languages. Further languages included English ($N = 12$), French ($N = 5$), and Turkish ($N = 1$). While the Egyptian sample therefore included the largest number of multilingual speakers, the Lithuanian sample included the largest range of different languages ($N = 10$). Interestingly, the languages were rather evenly distributed across the Lithuanian candidates, including nine speakers of Ukrainian and eight of Russian followed by two speakers of Albanian, Lithuanian, Kazakh, and Spanish respectively, alongside one speaker each for Afrikaans, Dari, English, and French. Overall, almost a quarter of the Lithuanian group reported speaking two languages.

In comparison, German was the most frequently spoken L1 in the Austrian sample with 61.5% of all participants. The remaining participants in Austria were L1 speakers of Italian ($N = 3$), English ($N = 1$) and Polish ($N = 1$). None of the Austrian test-takers reported to be a native speaker of a second or further language. As participants could indicate an unlimited number of L1s, some participants reported having up to four different L1s, and we need to acknowledge the possibility that participants might not have made a clear distinction between L1 and additional languages. Hence, the linguistic profiles of the students need to be interpreted with caution.

Thus, the sample included 14 students who considered themselves L1 users of English (but note the caveat in Table 7 below). Although this study is about contexts where English is not the majority language, we decided to keep these students in the sample, as they reflect real-world EMI settings. As outlined in the background section, we define EMI as “the use of the English language to teach academic subjects in countries or jurisdictions where the first language (L1) **of the majority of the population** is not English [emphasis added]” (Dearden, 2014). Although it is important to consider differences between L1 and L2 readers as discussed in the literature review, it is equally important to base studies such as the current one on real-world populations including L1 speakers of English, as omitting them (or treating them separately) would introduce bias in the results.

Table 7: Participants' languages as a percentage of the total according to the country of institution

Country of institution	Language*	N	%
Egypt (N=68)	Arabic	68	100
	English	12	17.6
	French	5	7.3
	Turkish	1	1.5
Lithuania (N=26)	Afrikaans	1	3.8
	Albanian	2	7.7
	Dari	1	3.8
	English	1	3.8
	French	1	3.8
	Kazakh	2	7.7
	Lithuanian	2	7.7
	Russian	9	34.6
	Spanish	2	7.7
	Ukrainian	11	42.3
Austria (N=13)	English	1	7.7
	German	8	61.5
	Italian	3	23.1
	Polish	1	7.7

*Participants were allowed to choose an unlimited number of “L1s” in the questionnaire. We therefore need to assume that students were equally proficient in the languages they indicated, but we cannot discern from the questionnaire responses whether students made a distinction between L1 and additional languages (except for those students who only indicated one language). This limitation is acknowledged by the authors.

Concerning the students' academic experience (see Table 8) and background (see Table 9), a considerable number of participants ($N = 88$; 82.2%) were in the first year of their bachelor's degree at the time of the data collection. In addition, the sample comprised three second- and three third-year students of bachelor programmes ($N = 6$). The Austrian candidates ($N = 13$) were all second-year students of their master's degree in ecology. In contrast, the Egyptian sample group was far more varied with regard to their field of study, containing students from a wide range of subjects, such as communication, political science, theology, or natural sciences, among others. Most Egyptian students, however, had a background in engineering (33.8%) and international business administration (26.5%). The Lithuanian candidates were relatively evenly distributed across different fields of the humanities and social sciences. Overall, the test-taker sample represents a relatively mixed group of academic disciplines.

Table 8: Participants' year of study for their intended major as a percentage of the total according to the country of institution

Country of institution	Year of study	N	%
Egypt	1	63	92.6
	2	2	2.9
	3	3	4.4
Lithuania	1	25	96.2
	2	1	3.8
Austria	2	13	100

Table 9: Participants' intended major as a percentage of the total according to the country of institution

Country of institution	Intended major	N	%
Egypt	Actuarial	1	1.5
	Biology	2	2.9
	Computer Engineering	1	1.5
	Computer Science	4	5.9
	Economics	4	5.9
	Engineering	23	33.8
	English Language and Literature	1	1.5
	Integrated Marketing Communication	4	5.9
	International Business Administration	18	26.5
	International Relations and Development	2	2.9
	Mass Communication	2	2.9
	Political Science	1	1.5
	Psychology	2	2.9
	Sociology	2	2.9
	Theology	1	1.5
Lithuania	Contemporary Communication	7	26.9
	English Language and Literature	5	19.2
	International Business Administration	9	34.6
	International Relations and Development	5	19.2
Austria	Ecology	13	100

3.1.5 Test administration

The reading test and the questionnaire were administered to groups of students in computer labs following standardized test administration guidelines. The guidelines outlined in detail the materials required for the administration and how the rooms and computers needed to be prepared prior to the test. The document also contained step-by-step instructions for test administrators on what to tell students before taking the test; for example, information on how to complete the test and submit their answers, or how to access and fill in the online questionnaire. Test administrators were advised on what to do if technical issues arose at the end of the document and were provided with an email address and phone number which were monitored by us during all the administrations. The test administration guidelines are included as Appendix B.

The test administration document was sent to the contact persons at the universities in Lithuania and Egypt. In addition to this document, the test administrators in Lithuania also asked us to give a brief training presentation about the test administration process. During this online presentation, the test administration document was discussed in detail, and we also demonstrated how to access and start the Aptis test, submit the test responses at the end, and how to complete the questionnaire. The presentation was recorded and uploaded onto a webserver so that all three test administrators in Lithuania could access it. The test administrators in Egypt felt that the test administration document was sufficient and that an online presentation was not necessary. For the EMI master's programme in Austria, we administered the test and questionnaire ourselves, following the standardized guidelines.

In line with the ethics requirements at our university, all students received an information document and consent form prior to the test administration. The information document briefly outlined the project and described what students were required to do. It also specified that all data collected would be kept confidential and stored securely. Students had to sign the consent form to take part in the study. The information document is included in Appendix C and the consent form in Appendix E.

After all candidates for one administration had arrived and been assigned a seat in the computer lab, the test administration commenced. Test administrators explained the procedure and handed out a keycode slip to each candidate, which contained a candidate reference number and a session code. Candidates had to enter this information in the online testing platform to start the test. Candidates were also told to note down their candidate reference number to be able to access their personalised results some weeks after the administration. Once they finished the reading test, candidates were instructed to raise their hand so that the test administrators could check whether the test responses had been submitted successfully. After finishing the reading test, each candidate was asked to fill in the online questionnaire. Candidates first had to enter their candidate reference number in the online questionnaire form so that their questionnaire responses could be linked to their test results. Once a candidate completed the questionnaire, they were free to go.

3.1.6 Reading test and questionnaire data analysis

The data was analyzed in several steps. First, we received the detailed test results and corresponding CEFR levels and Lexile reading measures from the British Council some weeks after each test administration. The data for all administrations was then combined. Second, the questionnaire data was combined and cleaned before linking it to the test results and Lexile score data through the candidate reference numbers. In a final step, the data was analyzed separately for each participating institution using different descriptive and inferential methods in SPSS (version 26 for Mac).

3.2 Lexile analysis of reading materials

As outlined above, we received reading materials from the participating EMI institutions to determine the texts' level of difficulty in terms of Lexile text measures to inform RQ1 and RQ2. However, before analysing the materials, we needed to ensure that the texts we used in our study were part of the course of all students who completed the reading test. In other words, texts which all or some of the students may not have encountered during their studies were excluded prior to the analyses, because such texts would not tell us anything about the direct interaction between text complexity and student proficiency. As the students from Egypt and Lithuania were recruited from a wide range of subject areas (see Table 9), we did not, therefore, 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). These core subjects were "Scientific Thinking" and "Philosophical Thinking" for the students in Egypt and "History of Western Civilization" and "Intercultural Communications" for the students in Lithuania. The core subjects were an integral part of all students' intended majors. Successful completion of the core subjects was compulsory and therefore important for the students' overall academic success. For the Austrian sample, who were all majoring in Ecology, we used subject-specific texts, which all students had encountered during their studies (mostly peer-reviewed journal articles).

Table 10 provides an overview of the texts we analyzed. 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 ($M=2974$), followed by the Lithuanian ($M=4274$ words) and Austrian texts ($M= 4406$ words). On the other hand, the Egyptian text sample also included the longest text of all ($Max = 22,846$ words), which is almost three times as long as the longest Austrian text ($Max = 8152$). Thus, as the large standard deviations suggest, all three EMI contexts used texts of varying length; yet, this trend was most pronounced in the Egyptian sample.

Table 10: 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

Once all texts had been collected and scanned, we sent them to MetaMetrics for analysis. It was necessary to convert some of the texts into .txt files prior to sending them, as the picture quality of the scans was not sufficient for the automated Lexile analysis. MetaMetrics then analyzed all texts and sent the Lexile text measures back to us via email some weeks later.

3.3 Interviews with lecturers

In addition to the data outlined above, we collected semi-structured interviews with lecturers of the participating students. Semi-structured interviews have the advantage of investigating a set of pre-defined questions with all interviewees while offering room for follow-up questions to gather additional information (Riazi, 2016). This procedure allowed us to probe the degree to which the lecturers felt that the students who attended their classes and thus worked with the texts we analyzed in this study were proficient enough in English to function in an EMI university setting, which informed RQ2. The lecturers used the reading materials we analyzed in their classes, which had the advantage that we also gained insights into how students need to work with the texts, as well as the texts' perceived level of complexity from the lecturer's perspective (RQ2).

In the following subsections, the interview recruitment procedure is outlined briefly first, followed by a description of the final interview participant sample. We then present the standardized guidelines and questions used in the interviews, and conclude with a discussion of the data analysis procedures.

3.3.1 Interview participant recruitment

The interview partners were recruited in collaboration with our contacts at the participating EMI universities. We asked our contacts to enquire at their institutions whether lecturers of participating students would be willing to take part in an online interview about the use of reading materials in class. The contacts then provided us with a list of potential interviewees, who were contacted by us directly. In total, we reached out to 10 lecturers across the three institutions via email, briefly outlining the study and asking whether they would be willing to take part. We received a positive reply from seven lecturers, who were then sent an information document (see Appendix D) and consent form (see Appendix E) by email, which they had to sign prior to participation.

3.3.2 Final interview participant sample

As described above, the final sample was made up of seven lecturers across the three EMI institutions. The interview participants were between 33 and 48 years old, with an average age of 42 (Table 11). The majority of lecturers (N=6) had at least 14 years of experience in teaching at university, with one participant having taught for 3 years. Similarly, most interviewees had taught EMI courses for a minimum of 14 years (N=5), while the two remaining participants had taught in English for 3 and 5 years, respectively. 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 Communications" (N=1). The Austrian interviewees were both lecturing various subjects in the Master's ecology course.

Table 11: Lecturers' age, experience, and subject areas

Country of institution	Age	Teaching experience (years)		Subject areas
		University	EMI courses	
Egypt (N=3)	46	14	14	Philosophy
	38	14	14	Scientific thinking
	33	3	3	Scientific thinking
Lithuania (N=2)	48	14	14	Theology, history
	44	19	20	Psychology
Austria (N=2)	44	17	17	Ecology, biology
	44	20	5	Ecology, science communication

3.3.3 Interview guidelines and procedure

The interviews with the lecturers were conducted following standardized guidelines (for the detailed guidelines, see Appendix F). A draft version of the interview was piloted with an external language testing researcher, and the guidelines were slightly adapted and revised based on the findings of this pilot. The final version of the interview guidelines consisted of six main thematic sections, which were based on the research aims.

After introductions, the first section explored how the lecturers select reading materials for their courses. We asked interviewees about 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, or whether they take student feedback into account when selecting the texts, among a number of other related questions.

The second section included questions about reading tasks. The participants were asked what their students typically need to do with the texts in the courses. The lecturers also outlined how they choose and assign the reading tasks and how students generally react to the tasks.

This was followed by a section exploring the types of support students receive when they struggle with text comprehension. Interviewees explained what they do if they realise that students find it hard to comprehend the texts, and whether they think their universities offer enough support to students.

The fourth section probed the lecturers' estimate of the reading materials' complexity. To that end, the interviewees were given five texts used in their (or their colleagues') courses, and they had to rank the texts from least complex to most complex. Lecturers were given 5 to 10 minutes to skim through and rank the texts, after which they were asked to give reasons for their rank order, i.e., why they ordered the texts in the way they did and which aspects they considered in their ranking.

In Section 5, the interviewees commented on their students' reading test results and associated Lexile reading measures. We briefly outlined the test results and explained how they compared with the Lexile analysis of the reading materials, after which the interviewees were asked whether the results match their teaching experience.

The final section of the interview included questions about potential repercussions of the current study on teaching practices. We asked lecturers whether the results of our study would affect their expectations towards students, and whether they thought that studies of this kind would be interesting for their department or institution.

The interviews were all conducted online via Jitsi or Zoom, depending on the participant's preference, and lasted about 45 to 60 minutes, depending on the length of answers from the participants. All interviews were conducted by two researchers (at different locations), with one researcher acting as the main interviewer. The second researcher monitored the interview guidelines throughout and asked additional questions which the main interviewer may have missed. Six of the seven interviews were conducted in English and one interview was conducted in German. For data backup purposes, both researchers independently recorded the interviews with audio and video recording software (QuickTime Player for Mac or Xbox Game Bar for Windows). After the interviews, the lecturers were sent a brief biodata questionnaire via email (see Appendix F), which they filled in and returned to us electronically.

3.3.4 Interview data analysis

To prepare the data analysis, the interview recordings were first automatically transcribed using the inbuilt MS Office 365 transcription software. We then checked the transcribed files and converted them to an .rtf-format for further processing. In the next step, one researcher listened to the entire audio file while adding time stamps using the transcription tool F4transkript Plus (Dresing & Pehl, n.d.). The researcher also checked the automatic transcription and corrected transcription mistakes. In addition, if the audio quality of a recording or passage was poor, the different backup files were consulted for clarification. For two interviews the quality of the automated transcription was inadequate, so these files were transcribed manually. Both automatic and manual transcriptions were adapted and formatted to fit the transcription guidelines for a simple transcription, as suggested by Dresing et al. (2015, pp. 27–30).

The data was then analyzed through *holistic thematic analysis*, as described by Holliday (2015). The data were first coded using MaxQDA 2020 (VERBI software, 2019) and the participants' comments were combined into larger codes organized around the interview questions. In a second step, the respective codes were further analyzed to identify recurrent subthemes across the participants' responses to one question. Whenever we identified recurrent themes within the individual questions or detected otherwise insightful comments, those statements were assigned subcodes to the questions they referred to. Upon completion of the coding process, extracts from the interviews were linked and used to exemplify the themes.

4. RESULTS

The results are outlined below following the two main research questions. RQ1 and RQ1a were informed by the students' test scores and equivalent Lexile reading measures, student questionnaire responses, and the Lexile text measures of the EMI reading materials. RQ2 and its four sub-questions drew on data from the student questionnaire survey and the interviews with the lecturers.

4.1 Research Question 1

RQ1: To what extent do EMI students' Aptis reading scores and equivalent Lexile reading measures match the demands of textual complexity that they will encounter in their EMI university context?

In the following section, the candidates' scores on the Aptis test and their corresponding CEFR and Lexile measures will be described first, before outlining the Lexile measures of the university reading materials. To answer the research question, the candidates' and texts' Lexile measures will be compared separately for each institution.

Table 12 below presents the candidates' results on the Aptis Reading Test, their corresponding CEFR levels, 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 master's students, the Egyptian and Lithuanian bachelor's 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 master's 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 12: 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

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

The Lexile text measures of the reading materials are displayed in Table 13, again separately for each institution. Data obtained through the Lexile analyses of the texts reflects the results on candidates' scores outlined above. First, the analyses identified a gap between the Egyptian/Lithuanian bachelor's context and the Austrian master's context when it comes to the reading materials the participants are assigned in their real-life university classes. The reading materials Lithuanian bachelor's 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 master's 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 ($p < .001$, $Z = -3.3$, $r = 0.6$).

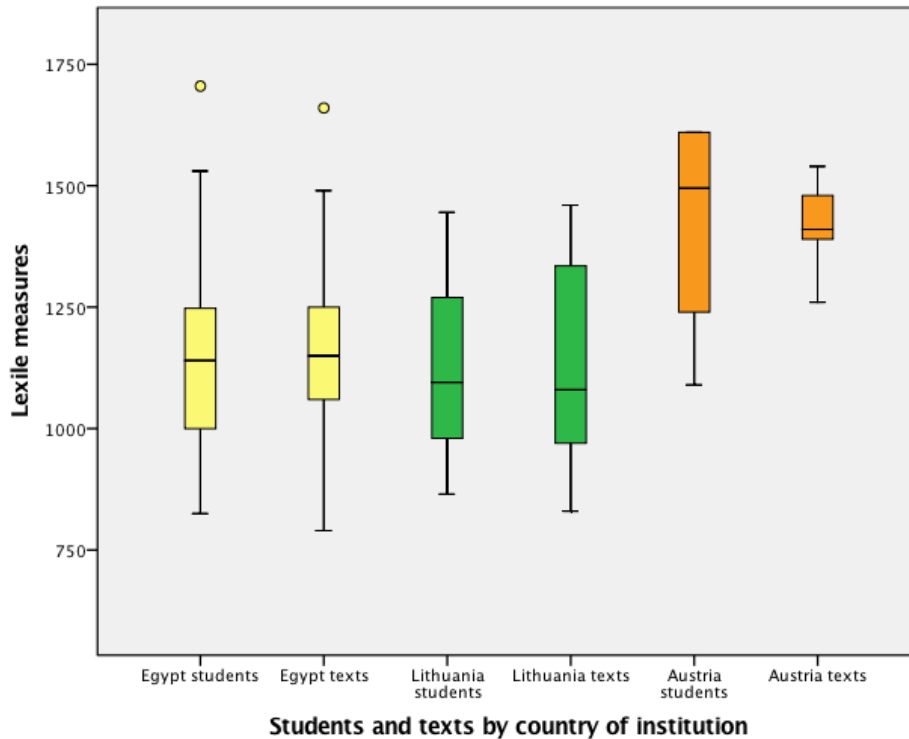
Second, within the Egyptian and Lithuanian samples, text complexity measures varied widely. The least challenging text linguistically (Lexile measure of 790L), as well as the most challenging text (1660L), were both identified in the Egyptian sample. Similarly, the Lexile measures of the Lithuanian sample 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 13: 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 ($M = 1119.3L$). Similarly, 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 ($M = 1412.3L$). The boxplot also illustrates a potential ceiling effect for the Austrian student sample due to the test used as discussed above.

Figure 1: Boxplot comparison of Lexile measures between students and texts across the three institutions



However, the lexical analyses also show that there are important differences in terms of Lexile measures between individual test-takers and individual texts (see Figure 2, Figure 3 and Figure 4). 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% 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% of 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% of students were placed at higher levels than the most difficult text (<1600L, see Figure 4) despite the above-discussed ceiling effect. In sum, while the average students' Lexile reading measures match the reading materials' average complexity relatively well, there are important 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

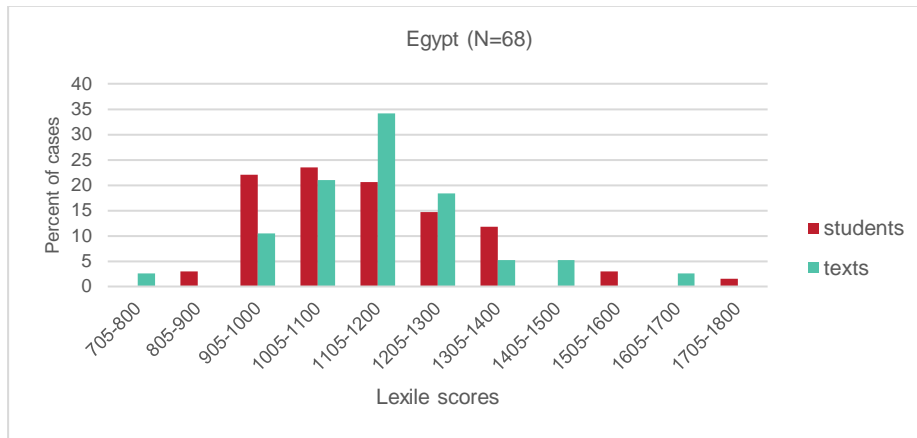


Figure 3: Comparison of Lexile measures between students and texts (% of cases at each level) for the institution in Lithuania

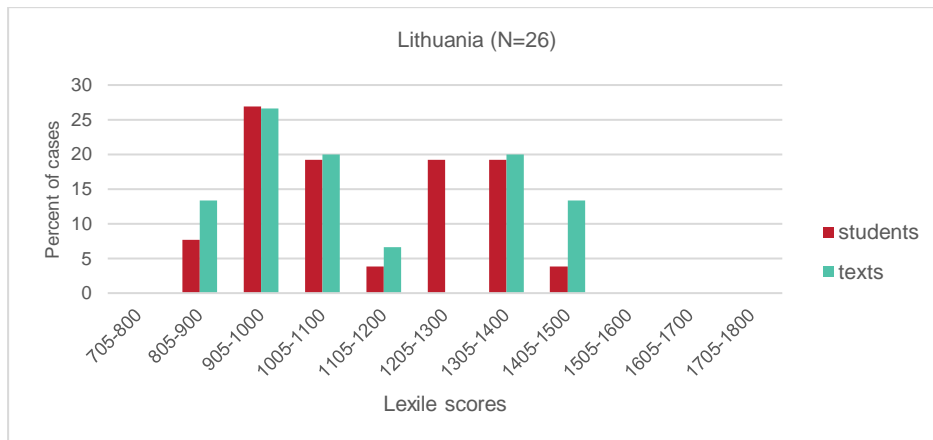
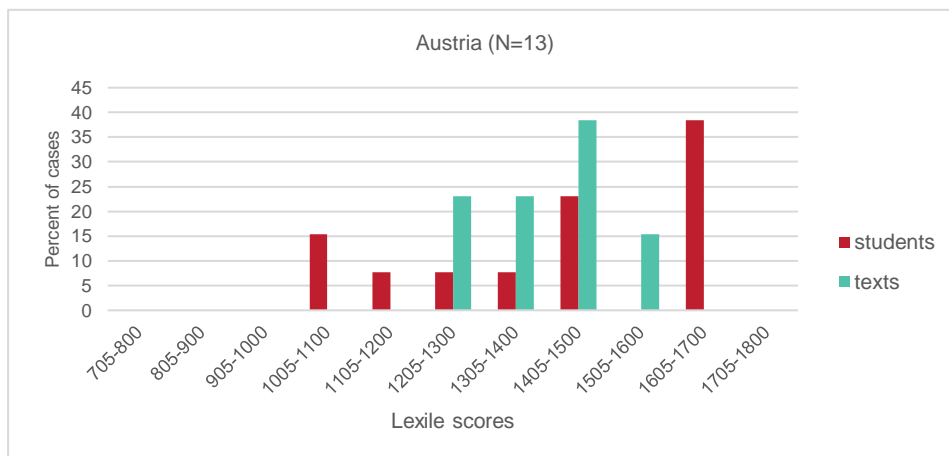


Figure 4: Comparison of Lexile measures between students and texts (% of cases at each level) for the institution in Austria



4.1.1 Research Question 1a

RQ1a: How similar do students think the texts in the Aptis reading tests are to the texts they need to understand in their EMI university courses?

This research question was investigated through student questionnaire responses. As shown in Table 14, when asked how similar the texts in the Aptis reading test were to the texts used in EMI university courses, the students (on average) felt that the texts were neither very similar nor very different. It is noteworthy that the Austrian master's students (who completed the Aptis General Test) perceived the Aptis texts to be more different compared to the Egyptian and Lithuanian bachelor's students (who completed the Aptis Advanced Tests). However, Mann-Whitney Tests revealed no statistically significant difference in students' perceptions between the three groups. It should also be noted that we could not directly compare the texts from Aptis tests with the texts used in the university settings for reasons of test security (students are randomly assigned different test versions with different texts, which means that we would have had to analyze a large number of texts currently used in live Aptis test versions).

Table 14: How similar did students think the texts in the Aptis reading tests were to the texts they need to understand in their EMI university courses?

Country of institution	Measure	
Egypt (N=68)	M*	3.3
	SD	1.2
	Min	1
	Max	5
Lithuania (N=26)	M*	3.3
	SD	1.0
	Min	1
	Max	5
Austria (N=13)	M*	3.9
	SD	1.4
	Min	1
	Max	5

*Students had to indicate their perceived level of similarity on a five-point Likert scale, where 1 was defined as "very similar", 5 was defined as "very different", and the intermediary categories (2, 3, and 4) remained undefined.

4.2 Research Question 2

RQ2 and its four sub-questions: If there is a mismatch between EMI students' scores and the textual demands in EMI courses as estimated by Lexile measures,

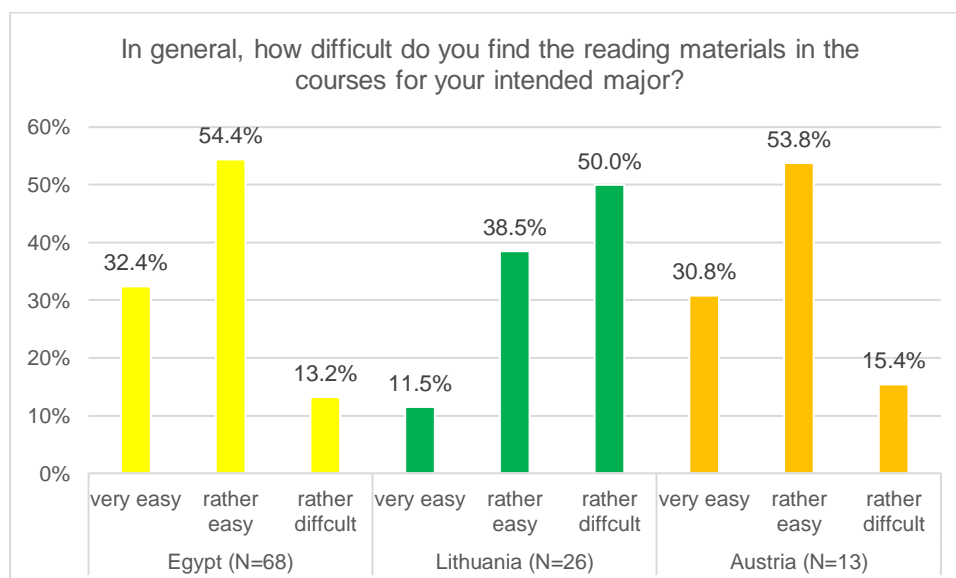
- a) **To what extent do students perceive the texts to be difficult?**
- b) **How do lecturers ask students to work with these texts?**
- c) **What kind of support students they get?**
- d) **How can course instructors explain this gap?**

Each of the four sub-questions will be answered in turn below, drawing on data from the students' questionnaire responses and the interviews with the lecturers.

4.2.1 To what extent do students perceive the texts to be difficult?

As shown in Figure 5, the majority of students across the three institutions stated that they found the texts "easy" or "rather easy", while a smaller number reported that they found the texts "rather difficult". However, none of the students thought that the texts were "very difficult" to understand. Considering the relatively large differences between individual students' proficiency and the texts' difficulty as presented in Figure 2, Figure 3 and Figure 4 above, one could have expected that a larger number of students thought the texts to be difficult. This is particularly true for the Egyptian sample, where about half of the students' proficiency (as expressed through Lexile reading measures) was below the level of complexity of about two-thirds of the texts (also as expressed through Lexile text measures, see discussion above). The responses from the Lithuanian and Austrian students better reflect the comparison between student proficiency and text complexity outlined above.

Figure 5: How difficult do students find the reading materials in the courses for their intended major?



Students who indicated that they found the reading materials in their EMI courses “rather difficult” were asked a follow-up question on why they thought the texts were difficult to comprehend. The results are displayed in Table 15. As shown, the most common reason for text being perceived as difficult by the Egyptian and Lithuanian bachelor’s students were “unfamiliar words and phrases”, followed by “unfamiliar topics and concepts”. Bachelor students from Egypt and Lithuania also thought that the “structure and organization of texts” was an important factor for text difficulty. Other reasons mentioned were the “length of texts” and “unfamiliar grammatical structures”. Results are similar for the two Austrian master’s students who struggled with the texts. In their replies, “unfamiliar words and phrases”, “unfamiliar topics and concepts”, and “length of texts” were all mentioned once as a reason.

Table 15: Why do students find the reading materials in the courses for their intended major difficult (multiple answers possible)?

Country of institution	Why difficult	N	%	% of cases
Egypt (N=9)	unfamiliar words and phrases	4	36.4	44.4
	unfamiliar grammatical structures	1	9.1	11.1
	unfamiliar topics and concepts	3	27.3	33.3
	length of texts	1	9.1	11.1
	structure and organization of texts	2	18.2	22.2
	Total	11	100	122.2
Lithuania (N=13)	unfamiliar words and phrases	9	37.5	69.2
	unfamiliar grammatical structures	2	8.3	15.4
	unfamiliar topics and concepts	5	20.8	38.5
	length of texts	4	16.7	30.8
	structure and organization of texts	4	16.7	30.8
	Total	24	100	184.6
Austria (N=2)	unfamiliar words and phrases	1	33.3	50
	unfamiliar topics and concepts	1	33.3	50
	length of texts	1	33.3	50
	Total	3	100	150

Similar trends emerged from the qualitative analysis of the interviews conducted with lecturers at each of the EMI institutions. First, when asked what aspects the lecturers considered in a text when estimating the level of difficulty for their students from a range of texts provided, all interviewees ($N = 7$) mentioned evaluating linguistic aspects. One concern expressed regarding language was vocabulary, referring both to the level of formality and the amount of subject-specific terminology.

When addressing this issue, two participants distinguished clearly between accessible language and specialised jargon, which they considered to be less easily accessible for students (Example 1 and Example 2), whereas interviewee I07 deemed technical terms particularly challenging for first-year students (Example 3). Furthermore, context-specific and archaic language were reported to be indicative of complex language (Example 4). Thus, not only students seem to predominantly report the importance of “unfamiliar words and phrases” in understanding a text, but the same feature is mirrored in the instructors’ reflections on text complexity.

- Example 1** *I06: And just yeah, sentence link, simplicity of some sentences and what words are they using? Are they using common everyday simple words that, you know, a young child could understand? Or maybe a teenager? Or are they using more advanced language? And also are they using technical terms?*
- Example 2** *I01: I do not know if this is the least complex, it seems to me Descartes because I think they will understand the words. Whereas the other ones are more technical and they will have to look up what all these words they /You know, as non-native speakers..*
- Example 3** *I07: So I think technical terms played a role because I assume that at freshman level, students maybe are not familiar with a lot of technical terms. Troposphere, CFC. (...) uh Isotope, right, uh?*
- Example 4** *I03: So not too much specific context-based historical or some other vocabulary, and the terms were defined. So what is plagiarism? How to organize? And so it was easy to / even if you didn't understand the term you would get it easily. As opposed to when there's a story like or the text there, and if you do not know what it is about, the only way you can guess it from the context or looking it up if you do not have the background/ like be difficult for you to understand the terms.*

Related to this, five out of seven interviewees reported that aspects of morphosyntax, in particular grammatical structures and sentence length, contributed to their judgement of text complexity. Similar to the students’ questionnaire responses, the lecturers also seemed to combine linguistic elements on both word and sentence level in their reasoning as to why a text might be more or less complex. Interviewee I02 (Example 5), for instance, stated that repetition in terms of sentence structure and parallelisms might make texts more easily accessible. Other responses linked text complexity to the length of phrases and sentences (Example 6), whereby their own educational background and experience as a non-native speaker of English served as a reference point (Example 7).

- Example 5** *I02: Yeah, good question. Yeah vocabulary. (...) Vocabulary and again, grammatical structure. Since it is largely it is a set of laws, simply. Once you read part of this (...) The grammar becomes or the / Well, conceptually becomes predictable. And grammatically, the sentences are fairly repetitive.*
- Example 6** *I05: I mean one is the terminology. The vocabulary used. And the second one would be the structure of the phrases if they are really complex and long phrases, or if it is shorter phrases. So if I had much more time now, I would go through and check the length of the sentences and the construction of the sentences. Yes.*

Example 7 *I04 [translated from German]: I actually tend to look at sentence length because if sentences get longer than those we are used to in German, then you get lost easily, in particular if you do not master the language that well. That is why I go for the one that has shorter sentences if I can choose between two articles that cover the same topic. In addition, I usually go for the article that is written in easier language based on my personal educational background.*

In addition, five out of seven interviewees reported that conceptual difficulty was a factor they took into account when judging text complexity. This again mirrors the self-reports by students, for whom “unfamiliar topics and concepts” was the second most important reason as to why they found reading materials difficult (see Table 15 above). Interviewee I06 explained in more detail what makes a text conceptually difficult to grasp (Example 8). For them, conceptual difficulty related to the number, density and presentation of ideas. Other lecturers reported that texts are conceptually more difficult if the students are not familiar with the topic (Example 9) or the text type (Example 10). Preparatory tasks such as pre-reading activities were considered by two informants to alleviate the challenge of unknown concepts and topics, perhaps because such preparation would also familiarize students with difficult vocabulary (Example 10; Example 11).

Example 8 *I06: So I would consider what kind of/ how many ideas are presented and how complex are those ideas. So do you need to know / is it a whole lot of information all at once? And how difficult are those concepts to understand? Is it something really simple like $2 + 2$ or is it something (...) more advanced that required a bit more steps to get to the conclusion if that makes sense?*

Example 9 *I03: Aha, first I looked at, you know, the topic and how often students you know encounter it. So we talk a lot about reading and writing and plagiarism. So just looking at the text, it is already easier because they deal with it every day at university. They hear it all the time and /*

Example 10 *I03: It was just difficult to read. I was trying to imagine how students would look at it and I guess they would read more texts four and five for the reading and writing classes. So they would be more familiar with that kind of writing and those kinds of text books than the other ones. So, for the others it would need pre-tasks and prepare them.*

Example 11 *I02: I am not judging the text as an outsider. I am judging the text, knowing the the pre-reading. The discussion that I give them so it is not quite fair. Probably if I I knew nothing about any of these texts, I would have ranked / The Code of Hammurabi is by far the most complex. Just because the language is so obscure, yeah.*

Apart from language and content variables, four participants also considered a number of other factors when estimating text complexity. Three respondents from the natural sciences attributed text complexity to certain writing styles of specific scientific journals and genres of scientific writing. Interviewee I04, for example, stated that depending on the academic journal, the authors were provided with more or less space to present their research, which in turn affects the density of information presented to the readers. The interviewee intimated that if readers are given additional context, text complexity decreases (Example 12).

Similarly, another lecturer in natural sciences identified direct quotations as an indicator for lower text complexity, drawing on his genre knowledge (Example 13).

Example 12 I04 [translated from German]: *As a result, some topics deal with/ as in the second article. That journal also allows longer articles, which means that authors are provided with slightly more time to introduce the topic. They have somewhat more space, not time, I mean space in the article to describe materials and methods and finally also assess it. That makes it, from my point of view, easier to read if the article does not only include the absolutely most important information but also allows to build a frame around. That is why I think, I mean this is based on my own assumption I need to admit, that it is easier for me to/ and that is similar to what I could observe with article No. 3. [...] And then there are the two Science articles (..) And in Science, it is about / there are only very short sentences. It is very, very good English. But each sentence contains a huge amount of information, which makes it tiring for someone to actually read it because it often takes you half an hour to read 2, 3, 4 sentences to grasp the entire context, including the literature that you might need to consult briefly, the literature that is cited and so on. That means it is quite exhausting a text to read because you progress only very slowly.*

Example 13 I05: *Yes, that was a little bit more of an intuition at the first text, Uh there were a lot of citations. [...] And actually in natural science that is not that common, however, in an editorial, or if you / Uh, yeah and and that is/ That was for me a little bit the sign, OK, that is written in an easier-going way. Let us say like more, how do you say colloquial?*

As in the student questionnaires, yet not as frequently, individual interviewees also regarded the structure, organization, and length of texts as factors for text complexity. One informant mentioned that the length of a text would affect how accessible it was perceived by students (Example 14), stating at a different point of the interview that experience had shown her that longer texts were relatively tiring for students. In addition, as Example 15 and Example 16 demonstrate, coherence and textual organization appear to have an impact on the readability of texts, with interviewees stating that clear sentence turns and lack of ambiguity, as well as clear organization and formatting, might make reading materials easier to comprehend for students.

Example 14 I07: *This is what I based my understanding on, and also because it is long.*

Example 15 I07: *There is no (..) Like a, you know, complicated / or turns in the sentence. So you need to a-/You will understand the meaning. There's no second or double meaning (..)*

Example 16 I02: *Because of the format, because it is chunked. Because it has an abundance of subheads.. It is organized. In terms of its organization, it is clear.*

4.2.2 How do lecturers ask students to work with these texts?

Text complexity measures do not include context-use factors, such as learners' background knowledge, the tasks students need to perform based on the text materials, and their embedding in a larger learning cycle. Hence, the interviews with lecturers provided complementary insights into how and why lecturers were using the texts we analyzed in class and what tasks the students needed to complete in relation to those texts. Lecturers were thus asked how and for which purposes they selected reading materials and tasks.

With regard to the mode of task administration, it emerged from the data that students were asked to perform reading activities in both oral and written form. While the majority of lecturers stated that they assigned written reading tasks, two lecturers explicitly mentioned assigning both written and oral tasks, and one lecturer focussed on the oral presentation of the written materials. One interviewee commented that the type of tasks they assigned had shifted towards more writing-based activities since the start of the COVID-19 pandemic due to increased remote teaching (Example 17). Another interviewee mentioned that the pandemic had led them to integrate new online-teaching tools to make oral activities work in a virtual teaching situation (Example 18). Overall, most interviewees mentioned that they assigned the reading tasks in both writing as well as orally to make sure their students understand what they are expected to do. However, one participant made clear that their way of assigning tasks had changed from an oral to a written format due to the pandemic.

Example 17 *I01: Now I will be focusing on, you know, commenting on the text. This is the main thing. So, because so I gave them a passage that I think is important. And then I asked them to explain, explain the main ideas, explain in their own words. A passage that is particularly important with, you know, important ideas or /and then I ask them to write a commentary on the text / as close as possible, and I usually give them an example in advance. Obviously, I write that one, you know, commentary, and then I post it, and then they'll do the same and this is / I think it is an important / I think it is just to make sure that they understand.*

Example 18 *I07: No. So, so when we had this emergency situation with COVID in the mid of last semester, I / To be honest, it wasn't really a good experience with me. Uh with online. Because I thought that the expectation from the university is that instead of teaching face-to-face I would teach virtually. So, some / most of the students did not attend the class. There was no clear policy from university regarding attendance. So, I decided to completely change my uh let us say strategy this time. This is why I use break-out rooms. I use it for the first time and it is working perfectly. And I have, I would say, 95% attendance rate, every single time.*

Most oral tasks mentioned by the lecturers were debates and presentations based on scientific topics, as the following examples illustrate. While interviewee I07 stated that they used reading materials to initiate debates for content learning purposes with undergraduate students (Example 19), the activities interviewees I04 and I05 mentioned from their postgraduate degree programme seem to be more directed towards scientific research. I05 describes that master's-level students are required to prepare a concept for their master's thesis, which they need to present in class (Example 20).

Similarly, in Example 21 students were asked to familiarize themselves with a topic and then present it to colleagues in their own words. Students were thus asked to transfer language from a written scientific register to spoken and less technical language.

Example 19 *I07: Yeah, we do debates. Like, sometimes I choose controversial topics. Like in-vitro fertilization, or ah I do not know uh cloning, you know. And they can read (..) at home and then use the information to debate each other. Some are pro; others are cons. Sometimes, I do activities in class, where I ask students to go into separate rooms online, okay? They debate in the class and they / obviously, I ask them questions to prepare among each other. So, they learn from each other. Right? And then, we do a class discussion. (..) These are the activities I mainly do.*

Example 20 *I05: Yes, they have the task to prepare a concept. Again, in both courses they prepare a concept they present in one of the courses. They present their project with a small presentation and then they work based on their content with the supervision of the lecturers.*

Example 21 *I04 [translated from German]: It is about recognizing the individual parts of a text: introduction, materials, methods, results and discussion. And it is about presenting the core of it, the most important points, to the colleagues (..) And the focus is quite clearly on passing the information on in a comprehensible manner. This means that students notice for themselves where they had experienced most difficulties when trying to understand and then, ideally, present those matters more comprehensibly than they might seem on the first glance when reading. So the work that was necessary to understand the content will be made easier for the others by providing a more comprehensible approach because they now know better how to formulate it to make it understood to someone who is at the same level of education as they are.*

The interviewees also mentioned a range of written tasks that students are required to perform based on the course reading materials. As with oral tasks, a recurrent activity was to make students reformulate the content of reading materials in their own words. In Example 22, interviewee I01 tried to apply a narrow focus in their assignments to ensure that students do not consult online sources instead of reading the original text. In addition, some lecturers reported to adapt the reading materials to the level of their students. As such, first-year university courses reportedly focused on basic academic writing skills, such as paraphrasing, summarizing and citing correctly (Example 23), or on acquiring and consolidating new knowledge (Example 24). At a more advanced level, in contrast, students worked more independently and were only supervised by the lecturers (Example 25).

Example 22 *I01: It obviously depends on the courses they are taking, um. But it is very important for me that they do not go to some site and just read what Socrates says in the Apology and stuff. For me, I won't be able to, you know, if I give them a different text without checking anything else in the online sources that they are able to read a philosophical text. And so I tried to make it as focused as possible. Otherwise, they'll go, you know, to SparkNotes or they'll go so consult something on just about what Socrates says. You know I am/ I said I am not interested in what the Internet says. I am interested in what you say about this and how you read this.*

Example 23 I03: *So mostly, I think about several questions, maybe for "Interpersonal communication" I just make them read the textbook and the tasks like this year I selected the key topics and so what / Because they are freshmen, so I selected key topics and they have to, in their own words, paraphrase and summarize and answer the questions. Like there might be a question: Compare what the difference is between interpersonal and impersonal communication. So in the textbook they have the version, but they have to make it their own. They have to write it and if there's something to cite, they have to cite it. So kind of making their own story and then when you should we give them an exercise where they have to apply the theory to their own life.*

Example 24 I02: *I give students note packets for each topic that we cover. So if we are dealing with the Roman Empire, I will give them a note packet that consists of a maybe three four pages and part of that note packet could be some questions that they need to answer from the textbook and part of it is my own lecture outline so it is as a paper with my outline on which they can take notes. I want them to take notes I think it is a good practice. That is a huge pedagogical question in itself, but I want them to take notes, and so I provide them help in doing so.*

Example 25 I05: *So actually all the courses are pretty advanced courses. So, and this compulsory module, it is like the final module they do to bring all their knowledge together. So they really do not get tasks like in a bachelor's course for example, but they work mostly independently with the supervision of the lecturers.*

Furthermore, subject-specific differences in how lecturers asked students to work with the reading materials emerged from the interview data. While humanities subjects seemed to concentrate more on fostering academic reading and writing skills as well as critical thinking by using text materials, lecturers in the natural sciences mentioned practice-oriented approaches more often, in that they asked students to apply the knowledge gained from the reading materials in case studies and experiments. For example, interviewee I03, an instructor in a humanities undergraduate program, asks students to read scientific studies and books for general understanding (Example 26), whereas I04's natural sciences students use the reading materials as a detailed guide for conducting experiments and case studies (Example 27).

Example 26 I03: *Sometimes I ask them to find several articles or books. Like if for practical part, they have to write about the field in psychology that they are interested in. So for instance, clinical or educational psychology, and so they have to research the field and we have to talk about what skills they have. What we would need to learn. So it is the reading is more kind of practical looking at / understanding what people in this field do and based on that they evaluate themselves and where they would like to be at. What skills they need to include and learn. So it is more maybe hands-on. So look, the reading is there to help them understand the field.*

Example 27 I04 [translated from German]: *That is a task in the seminar or in a lecture or a lecture-practical course. So, they take what they have read as a kind of recipe to do a practical implementation of the same based on information technology. Graphic information systems so to say or other tools for spatial modelling or mathematical modelling in general. So, you can understand it as a recipe that they try to rebuild, so to say.*

One of the interview questions was concerned with the role of reading and the reading amount in class. The lecturers were thus asked how much time students needed to spend on reading as a proportion of the overall time they spent on the course. The interviewees appeared to have difficulties in estimating the weekly reading workload for their students. Four out of seven participants estimated that reading the texts and doing the associated reading tasks might account for 20% to 30% of the overall time students spend on the class (I01 estimated that overall class work would take 4–6 hours a week). However, one interviewee thought that students might spend up to 80% of their time reading in particular courses. It is also noteworthy that interviewee I04, a lecturer in natural sciences, distinguished between the reading load in under- and postgraduate programs, with undergraduate programs potentially demanding more reading from students (Example 28). Overall, the results indicate that the amount of reading students are required to do depends on the subject area and the type of class.

Example 28 *[translated from German]*

Interviewer 2: Do students read less in their bachelor's programme compared to the classes they take at master's level, where the focus might be more on text reception, or what can you say about that?

I04: If we take the same course type into account, the master's programmes will include more reading than the bachelor's programmes. [...] However, we need to consider that there are more lectures for bachelor's students. Lectures as a course type, and there is generally more reading in lectures than in practical courses, such as lecture-practical courses or project studies.

Interviewer 2: So, this means the overall reading load / there is quite some reading for undergrads, too. Would you agree?

I04: If you consider the whole workload, there might be more reading in undergraduate programmes because it is about / undergraduate programmes are about building up a basis of knowledge, a basis that we can build up on later on by consolidating that knowledge and by putting theory into practice.

When it comes to the role of reading in class assessment, four out of seven interviewees agreed that although reading comprehension is necessary to understand the concepts taught, their course assessment does not explicitly test reading skills. The remaining three participants, however, stated that reading comprehension is to some extent required to successfully complete their course. Though they only play a little or minor role in the course assessment, I02 and I07 felt they need to include small quizzes in their weekly teaching to motivate their students to read the required texts. The majority of instructors thought the reading materials they provide serve as basic reading and thus introduction to a topic or concept. In addition, texts help to encourage critical thinking (I01), they serve as basis for discussion (I02), and are used for groupwork (I05, I07). Reading materials might also give additional background to consolidate content knowledge (I03), foster academic reading and working strategies (I04), and they might be a basis for project studies (I05, I06).

To sum up, this section has shown a variety of purposes and uses of reading materials in class. It highlighted the range of tasks students need to complete based on the text materials they are assigned, the lecturers' criteria for text as well as task selection, and the role of reading in class.

4.2.3 What kind of support do students get?

To answer this part of the research question, we again drew on both the student questionnaire responses and the interviews with the lecturers. Table 16 shows the results of the questionnaire responses on how students cope with difficult reading materials. As with the responses outlined in Table 15 above (Section 4.2.1), only students who indicated that they found the reading materials in their EMI courses difficult were asked this follow-up question. The responses to the questions were varied, but across the three institutions the most common type of support mentioned was “consult online resources”. A number of bachelor's students in the Egyptian and Lithuanian sample also stated that they “guess [the] meaning from context”, “consult a dictionary”, or “ask [their] course instructor” whenever they struggle to understand a text. Other coping strategies mentioned several times for these students were “ask student colleagues” or “look up resources in L1”. Almost half of the Lithuanian students who struggled with the reading materials also indicated that they translate the text into their L1, but this was only mentioned once by students from Egypt and Austria. Also, “take an English course” was not a common coping strategy for the students in the three institutions.

Table 16: How do students cope with difficult reading materials in the courses for their intended major (multiple answers possible)?

Country of institution	Coping strategies	N	%	% of cases
Egypt (N=9)	ask course instructor	4	19	44.4
	ask student colleagues	2	9.5	22.2
	use a dictionary	3	14.3	33.3
	consult online resources	4	19	44.4
	guess meaning from context	4	19	44.4
	translate into L1	1	4.8	11.1
	look up resources in L1	2	9.5	22.2
	take an English course	1	4.8	11.1
	Total	21	100	233.3
Lithuania (N=13)	ask course instructor	6	13.6	46.2
	ask student colleagues	5	11.4	38.5
	use a dictionary	7	15.9	53.8
	consult online resources	8	18.2	61.5
	guess meaning from context	8	18.2	61.5
	translate into L1	6	13.6	46.2
	look up resources in L1	4	9.1	30.8
	Total	44	100	338.5
Austria (N=2)	ask student colleagues	1	14.3	50
	consult online resources	2	28.6	100
	guess meaning from context	1	14.3	50
	translate into L1	1	14.3	50
	look up resources in L1	1	14.3	50
	take an English course	1	14.3	50
	Total	7	100	350

Data from the interviews provides further insights into students' coping strategies and institutional help from a different perspective. One way of coping with difficult reading materials, which features in both student questionnaires and lecturer interviews, is translating to the L1. Three out of the seven lecturers mentioned that some of their students seem to use machine translation tools such as *Google Translate* and *DeepL Translate*. The lecturers reported that students make use of the tools for reading and writing assignments (Example 29) or to get an overall idea about the content of a text (Example 30).

Example 29 *I06: So I will get one or two students that you can see they are typing things in Arabic and then running it through a translator. And then you are getting gibberish back, and they are not recognizing that the English bit that they are handing in to me is nonsense because they typed it into the translator.*

Example 30 *I04 [translated from German]: Because I believe that many students make use of these online tools, such as DeepL, or any other kind of translation tool. They copy passages from academic texts, paste them in the programme and then get a rough overview of the content, what is written in there. And I believe that many students work with these tools.*

When the lecturers were asked about student support, all instructors reported that they help students who struggle. Particularly in first-year courses, instructors aim to ensure that all students understand the reading materials by answering questions (Example 31) or giving them additional feedback (Example 32).

Example 31 *I02: Especially again because they are first-year students I try to do quite a bit of hand-holding as Americans are saying. Ensuring that they understand what was needed and answering the questions.*

Example 32 *I03: So I tried to give them a lot of feedback, especially for the first assignment. They plagiarize, they do all kinds of things, but for first one I just make a lot of comments, and ask them to re-do it so they learn.*

The interviews also revealed that student support at an institutional level differs among the three programmes included in our study. Although the Austrian EMI lecturers reported that there is an open-door policy and tutor system at their department if students need help, little institutionalized support is provided to students. This is mirrored by the fact that none of the Austrian students referred to the open-door policy and their lecturers as a potential help in the questionnaire. Instead, the necessary language skills seem to be acquired in a learning-by-doing approach, which one of the lecturers however deems to be sufficient. The students, on the other hand, seem to compensate for the lack of institutional support by helping each other (Example 33), and peer support was also mentioned in the student questionnaire responses (see Table 16 above). In addition, the second instructor at the Austrian university mentioned that the lecturers had taken the initiative to improve the current situation and offer a course on scientific writing, which may also help students in enhancing their reading skills (Example 34).

Example 33 I05: *There, to my point of view, there is no institutional support like above the courses. However, the [name of the degree programme] students are very well organized among each other. (..) So they help each other a lot. That is a really important aspect, but then within the courses (..) in both courses there are these lecturers which work as a supervisor. [...] If they have troubles, they can always contact us. They can contact us by email, or they come to the office. We do an appointment and we help them, but normally it is not language questions, so these are questions working on the topic. It is very rarely / I do not remember that I had questions regarding the language. No. And normally if / Yes, we had one or two students, Italian students, that were not that good at English, but then they helped each other.*

Example 34 I04 [translated from German]: *But in the master's programme, there is a compulsory module which is about presenting scientific content. It is about "scientific writing". In those courses, we actually explain the following: How can I read a scientific publication? How is a publication structured? And I think we integrated that in our curriculum on purpose to explain to our students how the format of an academic article works. Because it might be off-putting to read such a concise text.*

In comparison, the Egyptian EMI university has a more developed institutional network of student support to which the instructors regularly direct their students. The lecturers can refer their students to the "Center for Learning and Teaching" or the "Writing Centre" for language-related problems as well as further support in academic skills (Example 35). The lecturers appear to be satisfied with these support centres, although they lack feedback from the students (Example 36).

Example 35 I07: *So, anything related to linguistics and let us say English, the English aspect of the course. Let us say, something related to academic integrity or using Turnitin, I always refer them to either CLT, which is the "Center for Learning and Teaching" in [name of the university]. [...] I help them, I tell them, but I also emphasize that this is part of their English course as well.*

Example 36 I06: *I know that we have the "Center for Learning and Teaching", and my interactions with them have been always amazing. They are always really helpful. I do not have any feedback from students on their experiences. When they seek help. But based on my experience, I would think that they that they would have support.*

Similarly, the EMI institution in Lithuania offers student support at a number of levels. Interviewee I02 described the university-led tutoring system, where younger students get support from more advanced peers in a number of subjects and classes (Example 37). In addition, interviewee I03 mentioned that all students have an academic advisor, who is usually the first person to address if they face difficulties. The university staff also has an internal care system to identify students who may be struggling (Example 38). Such students are then referred to the "student support centre". If students do not meet the language requirements for their bachelor's course, the university provides the "PRIME" programme, which aims at enhancing language skills so that students can gradually integrate into Bachelor-level classes and ultimately join the regular bachelor's programme (Example 39).

In sum, the Lithuanian support system combines a range of offers for EMI students. This is also confirmed by students' questionnaire responses, which showed that a relatively large number of the Lithuanian students receive support from lecturers or peers (see Table 16 above). Despite this wide-ranging institutional support system, none of the students in this sample reported taking an English class.

Example 37 I02: *We have a tutoring centre. It is called the “Student Success Centre” but it is a tutoring centre. And that is largely upperclassmen. I mean we have staff member whose job it is to manage that. And that the director of that centre then has various upperclassmen who serve as tutors for everything from math to writing. Some tutors are kind of set up to help with / there are tutor students for specific classes. So that is something we make available [...]*

Example 38 I03: *Yeah, I think we are doing quite a lot, and if somebody needs more help, like you know, like we have also alerts. So we have a system. It is called “We Care System”. So if faculty member notices that a student is struggling with something they can talk to the student and report, you know, write a form what concerns they have and that goes to the student support centre and then, you know, they look into how to help the student. [...]*

Example 39 I02: *[name of the university] has a what they call a PRIME program and that is for students who come in and whose English is not sufficient really to do university-level work in [name of the university]'s view. So students who are in the prior “Prime” program P-R-I-M-E spend at least one semester completely in “Prime”. In the second semester of their prime year, if they have improved enough, some of them do a couple of courses, either one or two courses in the regular BA program.*

The lecturers in general seemed to be satisfied with the student support their universities provided on an institutional level. Two lecturers, however, voiced concerns that possible problems with students' (lack of) language skills might have a different cause. While interviewee I06 argued that entrance requirements should be revised to ensure that students are at the relevant level before entering the programme (Example 40), interviewee I04 thought that their colleagues expect too much of their students (Example 41).

Example 40 I06: *I think the issue might be that. The acceptance requirements, so when they are evaluating students competency when they are accepted. I think that might be the point where the problem rises / where you get students coming in aren't at the level that they are expected to be yet, but I do not get the impression that the university itself does not offer support when students are here.*

Example 41 I04 *[translated from German]: I think that students are well-prepared. I do not think that the university should offer additional courses or support. However, I sometimes wish my colleagues became more aware of how difficult it might be for a young high-school graduate to be suddenly working with an English-speaking workbook only. Of course, this is not a big deal for some who is a native speaker in English / I mean it is a big deal compared to those who experienced this transition period themselves. I think it is normal / [...]* This means we sometimes need to be more patient with your students.

To summarize, two out of three universities reportedly offer institutionalized help with language-related problems and further academic skills. The usefulness of these institutionalized support systems for improving reading skills, however, does not emerge from the interviews. Interestingly, the lecturers' observations and the support systems provided by the university only partly match with the students' self-reports on their coping strategies when encountering reading or language problems.

4.2.4 How can course instructors explain gaps between reading demands and reading abilities?

In the course of the interviews, we presented a summary of our analysis to the interviewees to gain further insights. Lecturers were asked to what extent the results match their personal observations as course instructors and what experiences they have had with text difficulty in their teaching. The reactions to the analysis were mixed, with four lecturers reporting that the results matched their observations entirely or to a large extent. Three other lecturers indicated that the results only partially matched their experiences. A number of interviewees were surprised that such a large number of students would have difficulty in understanding the course reading materials (Example 42).

Example 42 *I06: Yeah, so you have ones that are really good at the course and it is kind of a waste of their time because they are already / there's I'm not teaching them anything new and that should be the majority. So, the ones that are at the top end are way above everybody else. Then you have ones, wait, so you have a really wide spread of their ability and then you get the ones in the middle. And so it makes it challenging, to (...) reach all of that. To be on / you cannot be on the same level with all of them with your materials, so that having 50% of students not understand 2/3 of the, wait, 50% understood 2/3 of material. Is that right?*

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

I06: That would seem high. Over my course. (...) Uhm but I think that is because my course does not require a lot of reading comprehension. Some focus of it.

When asked about reasons, instructors recurrently referred to the large gap between the low- and high-performing students in their classes (as outlined in the extracts above), but a range of other explanations also emerged from the data. Interviewee I02, for example, mentioned that the reading gap between materials and student ability might be explained by the fact that the analyses did not consider what students do with the text. They said that some texts are used for illustrative purposes rather than for detailed understanding in their teaching (Example 43). Another possible reason for the reading gap mentioned by the lecturers was the students' range of dialects (Example 44). Interviewee I04, on the other hand, explained the relatively good match between texts and test-takers in the Austrian EMI master's programme was due to initiatives lecturers had taken in the past to increase the amount of English-speaking reading materials from undergraduate level onwards, leading to an increase in both language and academic skills (Example 45). Finally, interviewee I06 conceded that they do not proactively ask students for their feedback on reading materials, which might explain the mismatch between the lecturers' perspective and student ability (Example 46).

Example 43 I02: Yeah, I got the code of Hammurabi [text used at this institution] / what I the questions that I want that answer I am looking at it again here but (..) are um have to do with understanding something of the understanding of justice of that society. And also that so they grasp the hierarchical aspect. So I am using it really as in as an attempt to illustrate some things about (..) this ancient Mesopotamians society.

Example 44 I01: So um, I think it is a lot and. Yeah, it is just /, I think this might be the fact that it is different and also that even within the Arabic they have to know. Obviously the dialect that they grow up. They learn the dialect and you know.

Example 45 I04 [translated from German]: So we also believe that over the last couple of years, since we really started introducing English texts from BA-level onwards, but in form of text books. In the bachelor's programme, there are not many scientific articles yet. The students will mainly get teaching materials and learning materials in the English language and since we started to increase their amount in the BA programme, not exclusively, but we believe that we can start to see that master's students engage with English-speaking articles less unconfidently and indeed, the quality of presentations and the comprehension of content has (...) definitely been improved.

Example 46 Interviewer 1: OK. Um? Yeah, it is interesting. Have you experienced students expressing concerns about reading materials being too difficult or too easy for that matter in your classes?

I07: No. (..) In my class. So as I said, to be honest, I do not proactively ask for this feedback. Maybe this is something that I should do more often. Or usually I ask, how did you like the material? Both videos and reading, but I do not specifically focus on reading.

Interestingly, the interviews also revealed that the lecturers seemed to know little about the entrance requirements their students had to meet to be admitted for the EMI courses. Only four lecturers mentioned to be aware that their students' English language proficiency was tested prior to admission; however, they did know the details about this process (Examples 47, 48, 49 and 50).

Example 47 I01: I have not seen the language you know (..) exams that they have to take before you know, accepting a place at the enter [place of the university].

Example 48 I06: I do not know how they evaluate it. 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.

Example 49 I02: Yeah. I figured that TOEFL is good enough. They come in with taking that.

Interviewer 1: They have to take the TOEFL before they enter [name of the university]?

I02: Yeah before they come they have to. It is part of the requirements for university.

[...]: and I am pretty ignorant about that process. So you guys are the experts, but I mean, I am not sure what score would be required, but okay.

Example 50 *I05: I am not sure was it B-/ I think it is B1, but I am not totally sure about this, but I will check for you if you want, but you can send the application. I send you the/ if you want, I will send you the application site from the [name of the degree programme] course.*

Overall, course instructors' reactions to the results of our study were mixed and they provided individual explanations for their observations and impressions. There seems to be, however, a noteworthy lack of knowledge about admission criteria and language proficiency levels on the part of the lecturers.

5. DISCUSSION

In recent years, EMI degree programmes have steadily increased in academia across all subject areas due to internationalisation and the associated mobility of students across countries. Although EMI programmes usually require English language certifications from prospective students, we were not able to find research on how test scores obtained through large-scale international language examinations compare to the language demands of real-life EMI programmes outside of the UK/US context. In particular, previous literature paid little attention to the role of reading in EMI contexts, which is surprising as reading is deemed a core academic language skill (Aldridge, 2019; Au, 2000; Barth, 2003; Cox et al., 2003; Hermida, 2009; Schmeiser, 2009; Van Lanen et al., 2000). Prior readability studies were limited to the possible gap between L1 high school students' reading skills and the actual degree requirements in mainly L1 university settings. The current study is unique in that it investigated the gap between EMI students' reading proficiency and EMI university reading requirements across three different L2 contexts in Europe and Africa, using an innovative methodological approach.

We applied a mixed-methods design to gain new insights into text complexity in EMI university settings. Utilizing the Lexile Framework for Reading, measures of students' reading ability were compared with text complexity levels of authentic reading materials from three different EMI institutions. We then triangulated the reading proficiency and readability analyses with student questionnaire responses and semi-structured interviews with lecturers from the three institutions. The study appears to be the first to examine text difficulty not merely through quantitative text-based measures such as the Lexile measure, but also from the students' and lecturers' perspectives. Considering that these stakeholders are directly dealing with the texts on a daily basis, the findings from our study are arguably very relevant for practitioners in the field.

5.1 How do students' proficiency and measures of text complexity compare?

The first question this report aimed to answer was to what extent EMI students' reading proficiency, as measured through the Aptis General and Aptis Advanced reading tests, compares to text complexity of authentic EMI reading materials. Using the Lexile Framework as a reference scale, we found that average student proficiency and average text difficulty compare well for each of 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 a large number of texts, while other students would easily understand even the most complex texts. At the very least, a large number of texts would fall outside the recommended difficulty margin specified by MetaMetrics for a large number of students in these programmes.

It emerged from the student questionnaires that the texts from the Aptis reading suite used in the analysis seem to be reasonably comparable to the in-class reading materials. While some differences were expected due to the heterogeneity of student samples and academic subjects, students generally perceived the reading materials included in the Aptis test as neither very similar nor very different to the texts they read as part of their degree programme. As students were randomly assigned different forms of the same test, we cannot make any claims about the comparability of individual texts and why individual students perceived them as comparable to their usual in-class reading. Notably, however, the reading suite taken from the Aptis General Test, which was used in the Austrian master's-student sample, was reported to be more different, albeit not considerably different, from students' in-class reading materials.

The interviews with lecturers revealed insights on why the complexity of reading materials might deviate from students' level of proficiency. First, the data indicates that not all lecturers notice whether their students are proficient enough to deal with the complexity of their reading materials. Also, the lecturers showed a lack of knowledge about language entrance tests. Furthering lecturers' assessment literacy might therefore contribute to a better understanding of their students' language level at the beginning of an EMI university programme. This could, in turn, potentially also help lecturers judge the difficulty of reading materials, as several participants reported that they struggled with estimating text complexity and generally have to rely on their experience, rather than objective external criteria.

Furthermore, the observed range in student proficiency confirms findings by Lam and Maiworm (2014), who identified a large gap in proficiency between the highest- and the lowest-performing students in EMI programmes. In the current study, the heterogeneous sample may be attributed to a number of causes. First, although all the students were pursuing a degree in an EMI university programme, students came from a wide variety of L1 contexts and educational backgrounds. Each EMI institution thus came with its particularities and was embedded in local contexts. While the Austrian master's programme was part of a public university, in Egypt, for example, EMI institutions are private.

Another possible reason for the range in student proficiency may have to do with the language tests used for admission. Although it is not fully clear which language tests are accepted for admission across the three institutions, the interviews with lecturers showed that some admission tests might lack standardization or may not be adequate tools to discriminate between students with the necessary reading skills from those whose English is insufficient for the demands of the degree programme. In other words, the type of reading tested in the language admission tests may, to some extent, be incompatible with the actual reading demands of the EMI programmes. This mirrors 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 require localization and adaptations to authentically measure academic reading demands of EMI programmes.

Similar to the large range in student proficiency, our analyses also showed that texts used in EMI teaching varied widely with regards to complexity as measured through Lexile text measures. This was more pronounced for EMI bachelor's courses, where texts ranged between 790L and 1660L. The range of text complexity levels identified in the current study is thus greater than in comparable studies in an English L1 university context, where reading materials are generally more homogeneous in terms of text complexity (Wilkins et al., 2010; Williamson, 2004; Williamson et al., 2016).

Despite this large spread in text complexity, the *average* Lexile text measures of EMI reading materials in our study are similar to Lexile text measures found in previous research in an L1 context. Texts for beginner students of the EMI bachelor's 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 US (Wilkins et al., 2010, who reports a median of 1144L), and to grade 11 and grade 12 texts in the US (Williamson et al., who report a median of 1130L). Koons et al. (2016) report a slightly lower median of 1030L for key stage 4 texts in the UK. In comparison, Williamson (2004) reports median Lexile text measures of 1355L for first and second year L1 US university texts combined, thus indicating that text complexity levels rise throughout the first two years of university. Correspondingly, text complexity in the Austrian master's programme was higher than in the undergraduate programmes, with an average of around 1420L. Thus, EMI master's-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).

5.2 What makes texts difficult?

Our data analyses identified five main reasons for text difficulty in EMI university reading. The most common reason mentioned frequently by both students and lecturers is unfamiliar vocabulary. 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 also corroborate vocabulary frequency as one of the main underlying factors in calculating readability scores such as the Lexile Framework (see also DuBay, 2004).

The second factor influencing scores on the Lexile Framework is sentence length, and this was also mentioned as one reason for text difficulty in the data collected for the current study. Lecturers' comments indicated that texts consisting of long and complex sentences may be more difficult to comprehend for students. Students themselves did not directly mention long sentences as a factor for text difficulty (perhaps because it was not included as a pre-defined category in the student questionnaire).

Apart from lexical difficulty and sentence length, our study identified three additional factors which play a role in determining a text's readability. These factors, however, are not considered for calculating readability indices such as Lexile text measures. One of them is the students' familiarity with topics and concepts, which was the second most important factor for text difficulty mentioned by students who struggled with the texts. Lecturers also frequently referred to topic familiarity, as well as conceptual clues within a text, when asked about what makes texts difficult to comprehend for students. These results thus confirm work by Carrell (1987), who found that available content schemata are a critical factor in ESL reading comprehension success. Based on these findings, it seems important to refer back to DuBay (2004), who argued that readability indices like the Lexile Framework 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 considered to be 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). The Lexile Framework, however, does not seem to consider text length in its readability score.

Finally, another factor which seems to influence text complexity, but which is unaccounted for in readability measures such as the Lexile Framework, is the texts' structure and organization. Bachelor's students who struggled with the texts 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. Future revisions of the Lexile Framework could thus try to incorporate a text's coherence and cohesion for calculating text difficulty, as has also been highlighted in research by Crossley and colleagues (Crossley et al., 2007, 2008, 2011, 2017).

5.3 Which reading tasks are EMI students asked to do?

To contextualize the reading materials within the EMI university settings, we went beyond automated readability analyses and also examined what types of tasks students are asked to do based on the texts they read as part of their degree programmes. Studying reading tasks is important for understanding the level of comprehension required from students, as task demands influence the products and processes of reading (Alderson et al., 2015). For example, texts might be perceived more or less challenging depending on whether students need to showcase detailed understanding of a text as part of a reading task, or whether it suffices for them to grasp the text's overall topic.

Even proficient readers might face difficulties if a task requires higher-level reading processes or if the purpose of reading is unclear (Alderson et al., 2015). On the other hand, if students only need to grasp the topic of discussion at surface level, they may still be able to accomplish the task even if they do not understand all words in the text.

As the Lexile score is designed to measure text complexity of reading materials only, based on lexical and syntactical measures, it does not take the subsequent task into account. In teaching, however, reading materials are frequently accompanied by learning tasks. The interviews revealed there are a wide variety of both oral and written tasks that students need to complete based on the reading materials. Lecturers mentioned that students need to read texts and briefly summarise them in their own words, read texts and then debate the overall topic in class, use texts for presentations and concepts for dissertation theses, or work with texts independently for a specific project. The type of task also seems to vary depending on the subject areas students study. Some of these tasks arguably demand a more complete and detailed understanding from students (e.g., using a text to derive a concept for a dissertation thesis, or to replicate an existing research study), while for other tasks students may only need to understand the overall meaning but not necessarily all the details (e.g., for debating the topic in class, or when asked to read for general understanding of a concept). Hence, course instructors might have mitigated text complexity by balancing the demands of either text and task for their respective teaching goal and target group. To some extent this may also help explain why we found such a wide range of texts in terms of complexity as indicated through Lexile text measures or why lecturers' judgements of text complexity differed from Lexile text measures.

For these reasons, it is important to keep in mind that comparisons of a text's Lexile level with a student's Lexile reading measure do not take into account the complexity of the task that the student needs to perform based on the text.

5.4 What kind of support do students get?

Student support differed widely among the three EMI institutions. The interviews with lecturers in Egypt and Lithuania revealed an established support system at their institutions for students who struggle with English. However, students in Austria depend on tutors, lecturers, and (above all else) peers to help them with language-related issues. This difference may again be related to the type of degree programmes we investigated: While participants from Egypt and Lithuania were first-year EMI university students studying for a bachelor's degree, participants from Austria were in their second year of a master's programme.

Still, 2 of the 13 Austrian master's students indicated that they find the English reading materials for their EMI courses difficult. Interestingly, only one of them achieved a Lexile score which was below the Lexile score of all the texts analyzed for this student sample. The other student scored highly on the Aptis test and should, in theory, be able to easily understand most of the reading materials. According to the questionnaire data, this student gave "unfamiliar words and phrases" as well as "unfamiliar topics and concepts" as reasons for finding the text difficult.

Conversely, there were several students across the three institutions who did not find the texts difficult, although they achieved considerably lower Lexile reading measures compared to the texts' Lexile levels. Although it is not clear from the questionnaire data why these students did not seem to struggle with the texts, the interviews with lecturers indicate that it could have to do with the tasks students needed to perform. In other words, even though the texts' Lexile level was higher than students' Lexile reading measures, students did not struggle to complete the reading tasks, perhaps because the tasks did not require detailed comprehension. However, they also highlight that readability indices such as Lexile measures appear to be only a relatively coarse indicator for how well students can cope with a text. Our results indicate that individual differences between students, such as anxiety, self-confidence or motivation, as well as the tasks students need to perform based on the texts, also seem to play a role for text comprehension.

6. LIMITATIONS OF THE CURRENT STUDY

Even though this study has furthered our understanding of reading demands in EMI contexts, the generalizability of the findings is subject to a number of limitations. First, the scope of this study is limited by a relatively small sample size, especially with regards to the student sample. Although the student sample as a whole was relatively large ($N = 107$), the individual sub-groups were smaller, particularly the Lithuanian and Austrian samples.

A related shortcoming of the study is the heterogeneity of student cohorts and EMI backgrounds. Not only were participants educated in different geographic settings, but they emerged from different academic fields. Hence, text materials also included a range of subjects and differed in length, and the influence of subject-specific trends and different scholarly traditions cannot be excluded. Nevertheless, including students from different EMI contexts also allowed us to compare reading demands among institutions. In addition, it seems important to include currently underrepresented geographical regions in research on EMI teaching. This is particularly true for regions across Africa, where the challenges of EMI teaching are arguably greater than in other parts of the world.

A further 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, in their responses to more general questions, it cannot be excluded that lecturers also referred to classes and texts other than the ones analyzed in this study.

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 offers 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 which 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.

Although this was not the main aim of the study, another weakness was that we were not able to directly compare texts from the Aptis test suites with EMI university reading materials to answer research question RQ1a but had to rely on student questionnaire responses instead. Additional data on readability features in reading tests, as well as on authentic EMI reading materials and how they compare, are necessary to better understand reading test scores.

7. AREAS FOR FUTURE RESEARCH

There are two main areas which would benefit from further research based on the findings of the current study. First, similar to Green et al. (2010), future studies could encompass a more fine-grained analysis of textual features to explain a potential mismatch between students' Lexile reading measures and Lexile text measures of reading materials. This seems necessary, as readability indices such as Lexile only take a limited number of textual features into account (i.e., word frequency and sentence length). As research by Crossley and colleagues has also shown, it can be insightful to also include additional features from psycholinguistic research, as well as discourse and conceptual analysis, to achieve more accurate classifications, particularly for an L2 population. Including factors of textual coherence and variables such as lexical coreferentiality or syntactic sentence similarity, alongside word frequency (e.g., by using automated tools such as Coh-Metrix), appears to be a promising approach (Crossley et al., 2007, 2008, 2011, 2017).

However, while including additional quantitative analysis of textual variables can be insightful, the current study has highlighted that it is crucial to triangulate results with qualitative data. We were able to show that understanding the type of tasks students need to perform based on the texts, as well as students' topical and conceptual knowledge, also seem to be important factors for shedding light on the complex link between text complexity and student proficiency. In addition, individual differences between students also seem to play a role (see also Li, Hiver, & Papi, 2022).

Future studies should thus also investigate how all of these factors interact, for example, through mixed effects modelling, to further our understanding in this area. In addition, future research could compare EMI settings as defined in our study with programs delivered to L2 English speakers in L1 English countries.

Finally, based on our experiences during data collection, we would strongly encourage researchers to obtain agreement-in-principle from institutions to participate before proposing a study of this kind to the funder. Data collection for the current study suffered from repeated setbacks, as most institutions we contacted did not reply or could not take part for a variety of reasons.

8. CONCLUSION

This report set out to compare EMI university reading materials with L2 students' reading proficiency based on the Lexile Framework for Reading. The study found that on average, students' language proficiency and reading demands in EMI university courses match relatively well. We also found that readability measures of texts used in EMI university teaching are similar to measures of texts used in comparable L1 university contexts. However, the analyses demonstrated that EMI universities face a very heterogeneous student population in terms of reading skills. Similarly, reading materials in EMI degree programmes vary widely with regards to text complexity, more so than reading materials in L1 postsecondary reading.

The second major finding is that readability scores, although useful as general indicators for text complexity, should ideally be accompanied by additional analyses to more fully understand the extent to which students can cope with texts. By triangulating the lexico-syntactic analyses with data from student questionnaire responses, as well as interviews with lecturers, this study identified five main causes for text difficulty: (1) unfamiliar vocabulary; (2) unfamiliar topics and concepts; (3) long and complex sentences; (4) text length; and (5) unclear structure and organization of texts. Given that automated readability analyses such as the Lexile Framework only include measures of vocabulary frequency and sentence length, they do not appear to offer a complete picture of text complexity for a specific EMI student population. This finding was further exacerbated by the fact that students need to display different levels of comprehension depending on the task they need to perform, and they may also display different processes due to individual difference factors such as anxiety.

Future research could thus focus on the interaction between different potential variables of text difficulty through methods such as mixed effects modelling to help refine current readability measures. This would not only be to the benefit of lecturers but also, importantly, to students at EMI institutions around the world.

REFERENCES

- Alderson, J. C., Haapakangas, E.-L., Huhta, A., Nieminen, L., & Ullakonoja, R. (2015). *The diagnosis of reading in second or foreign language*. Taylor & Francis.
- Aldridge, D. (2019). Reading, engagement and higher education. *Higher Education Research & Development*, 38(1), 38–50. <https://doi.org/10.1080/07294360.2018.1534804>
- Au, K. H. (2000). A multicultural perspective on policies for improving literacy achievement: Equity and excellence. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research (Vol. III)* (pp. 835–851). Lawrence Erlbaum Associates.
- Austrian Federal Ministry of Education Science and Research. (2018). *University report 2017*.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford University Press.
- Barth, P. (2003). A new core curriculum for all: Aiming high for other people's children. *Thinking K–16: A Publication of The Education Trust*, 7(1), 3–31. <https://files.eric.ed.gov/fulltext/ED476116.pdf>
- Beinborn, L., Zesch, T., & Gurevych, I. (2014). Readability for foreign language learning: The importance of cognates. *ITL – International Journal of Applied Linguistics*, 165, 136–162. <https://doi.org/10.1075/itl.165.2.02bei>
- Biber, D., Conrad, S. M., Reppen, R., Byrd, P., Helt, M., Clark, V., Cortes, V., Csomay, E., & Urzua, A. (2004). Representing language use in the university: Analysis of the TOEFL 2000 spoken and written academic language corpus. *TOEFL Monograph Series*, January. <http://www.ets.org/Media/Research/pdf/RM-04-03.pdf>
- Brenn-White, M., & Faethe, E. (2013). *English-taught master's programs in Europe: A 2013 update*. https://www.wittenborg.eu/sites/default/files/news_pdf_files/English-Language-Masters-2013-Update.pdf
- British Council. (2020a). *Aptis*. <https://www.britishcouncil.ge/en/exam/aptis/what>
- British Council. (2020b). *Aptis Advanced*. <https://www.britishcouncil.es/en/exam/aptis/advanced>
- British Council. (2020c). *Aptis Advanced technical supplement version 1.0: Technical report*. https://www.britishcouncil.org/sites/default/files/aptis_advanced_technical_supplement_final.pdf
- Carlsen, C. H. (2018). The edequacy of the B2 Level as university entrance requirement. *Language Assessment Quarterly*, 15(1), 75–89. <https://doi.org/10.1080/15434303.2017.1405962>
- Carrell, P. (1987). Readability in ESL. *Reading in a Foreign Language*, 4, 21–40.
- Carroll, J. B., Davies, P., & Richman, B. (1971). *The American heritage intermediate corpus*. American Heritage Publishing Co.
- Chall, J. S., & Dale, E. (1995). *Readability revisited: The new Dale-Chall readability formula*. Brookline Books. <https://doi.org/10.1080/0163853X.2017.1296264>
- Clifford, V., & Montgomery, C. (2011). *Moving towards internationalization of the curriculum for global citizenship in higher education*. The Oxford Centre for Staff and Learning Development.
- Coleman, J. A. (2006). English-medium teaching in European higher education. *Language Teaching*, 39(1), 1–14. <https://doi.org/10.1017/S026144480600320X>

- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. CUP. <https://rm.coe.int/1680459f97>
- Cox, S. R., Friesner, D. L., & Khayum, M. (2003). Do reading skills courses help underprepared readers achieve academic success in college? *Journal of College Reading and Learning*, 33(2), 170–196. <https://doi.org/10.1080/10790195.2003.10850147>
- 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., Louwerse, M., McCarthy, P., & McNamara, D. (2007). A linguistic analysis of simplified and authentic texts. *The Modern Language Journal*, 91, 15–30. <https://doi.org/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>
- Dale, E., & Chall, J. S. (1949). The concept of readability. *Elementary English*, 26(1), 19–26. <http://www.jstor.org/stable/41383594>
- Dearden, J. (2014). English as a medium of instruction: A growing global phenomenon. In *English language and higher education*. https://www.britishcouncil.es/sites/default/files/british_council_english_as_a_medium_of_instruction.pdf
- Dresing, T., & Pehl, T. (n.d.). *F4transkript Plus*. dr. dresing & pehl GmbH.
- Dresing, T., Pehl, T., & Schmiederer, C. (2015). *Manual (on) transcription: Transcription conventions, software guides and practical hints for qualitative researchers* (3rd Englis). https://www.audiotranskription.de/download/manual_on_transcription.pdf
- DuBay, W. (2004). *Principles of readability*. Impact Information. <http://www.impact-information.com/impactinfo/readability02.pdf>
- Dunn, K. (2020). *Aptis scoring system Version 2.0. Technical report TR/2020/002*. https://www.britishcouncil.org/sites/default/files/aptis_scoring_system_v2.0.pdf
- Educational Testing Service. (2020). *TOEFL ITP® overall performance descriptors*. https://www.ets.org/toefl_itp/research/performance-descriptors/
- 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. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52008DC0566&from=EN>
- 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. https://www.lemmens.de/dateien/medien/buecher-ebooks/aca/2014_english_taught.pdf
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32, 221–233. <https://doi.org/10.1037/h0057532>

- Frank, A., Hieronimus, S., Killius, N., & Meyer-Guckel, V. (2010). *Rolle und Zukunft privater Hochschulen in Deutschland: Eine Studie in Kooperation mit McKinsey & Company* (p. 98). Stifterverband für die deutsche Wissenschaft. <https://www.stifterverband.org/download/file/fid/269>
- Fry, E. B. (1968). A readability formula that saves time. *Journal of Reading*, 11, 513–516. <https://www.jstor.org/stable/pdf/40013635.pdf>
- 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). [https://www.teachingenglish.org.uk/sites/teacheng/files/H035 ELTRA Internationalisation_HE_and the growing demand for English A4_FINAL_WEB.pdf](https://www.teachingenglish.org.uk/sites/teacheng/files/H035%20ELTRA%20Internationalisation_HE_and%20the%20growing%20demand%20for%20English%20A4_FINAL_WEB.pdf)
- 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., Ünal, 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>
- Gunning, R. (1952). *The technique of clear writing*. McGraw-Hill.
- Gürtler, K., & Kronewald, E. (2015). Internationalization and English-medium instruction in German higher education. In D. Slobodanka, A. K. Hultgren, & C. Jensen (Eds.), *English-medium instruction in European higher education. English in Europe, volume 3* (pp. 89–118). Walter de Gruyter.
- Hermida, J. (2009). The importance of teaching academic reading skills in first-year university courses. *The International Journal of Research and Review*, 20–30. <https://www.mansfield.edu/fye/upload/Academic-Reading-Skills.pdf>
- Holliday, A. (2015). Qualitative research and analysis. In B. Paltridge & A. Phakiti (Eds.), *Research methods in applied linguistics: A practical resource* (pp. 49–62). Bloomsbury Publishing. https://books.google.at/books?id=wU69CAAQBAJ&printsec=frontcover&hl=de&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false
- 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>
- IELTS. (2020). *Common European Framework*. <https://www.ielts.org/ielts-for-organisations/common-european-framework>
- Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel. In *CNTECHTRA Research Branch Report*. <https://stars.library.ucf.edu/cgi/viewcontent.cgi?article=1055&context=istlibrary>
- Koons, H., Bush, L., & Sanford-Moore, E. E. (2016). *Primary and secondary textbook complexity in England*. <https://metametricsinc.com/research-publications/primary-secondary-textbook-complexity-england/>
- Lam, Q. K. H. 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. https://www.lemmens.de/dateien/medien/buecher-ebooks/aca/2014_english_taught.pdf
- Li, S., Hiver, P., & Papi, M. (2022). *The Routledge handbook of second language acquisition and individual differences*. Routledge.

- 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>
- Maiworm, F., & Wächter, B. (2014). Part I – The big picture. In B. Wächter & F. Maiworm (Eds.), *English-taught programmes in European higher education: the state of play in 2014* (pp. 25–65). Lemmens Medien. https://www.lemmens.de/dateien/medien/buecher-ebooks/aca/2014_english_taught.pdf
- Manarin, K., Carey, M., Rathburn, M., & Ryland, G. (2015). *Critical reading in higher education: academic goals and social engagement*. Indiana University Press.
- McCormick, A. (1999). Credit production and progress toward the bachelor's degree: An analysis of postsecondary transcripts for beginning students at 4-year institutions. *Postsecondary Education Descriptive Analysis Reports*, 179, 1–65. <https://nces.ed.gov/pubs99/1999179.pdf>
- McIlwraith, H., & Fortune, A. (2016). *English Language Teaching and Learning in Egypt: an insight*. British Council.
- Mesmer, H. A. E. (2008). *Tools for matching readers to texts: Research-based practices*. Guilford Press.
- MetaMetrics. (2020a). *About Lexile® Measures for Reading (for education companies & publishers)*. <https://metametricsinc.com/education-companies/lexile-for-reading-and-writing/about-lexile-for-reading/>
- MetaMetrics. (2020b). *About Lexile® Measures for Reading (for educators)*. <https://lexile.com/educators/understanding-lexile-measures/about-lexile-measures-for-reading/>
- MetaMetrics. (2020c). *About Lexile® Measures for Reading (for parents & students)*. <https://lexile.com/parents-students/understanding-your-lexile-measure/lexile-measures-reading/>
- Mezek, S. (2013). *Advanced second-language reading and vocabulary learning in the parallel-language university*. Stockholm University.
- O'Sullivan, B. (2015). *Linking the Aptis reporting scales to the CEFR. Technical report TR/2015/003*. British Council. https://www.britishcouncil.org/sites/default/files/tech_003_barry_osullivan_linking_aptis_v4_single_pages_0.pdf
- O'Sullivan, B., Dunlea, J., Spiby, R., Westbrook, C., & Dunn, K. (2020). *Aptis General technical manual version 2.2: Technical report*. British Council. https://www.britishcouncil.org/sites/default/files/aptis_technical_manual_v_2.2_final.pdf
- Owen, N., Shrestha, P., Hultgren, A. K. (2021). *Researching academic reading in two contrasting English as a medium of instruction contexts at university level*. ETS Research Report No. RR —21-04. <https://doi.org/10.1002/ets2.12317>
- Papageorgiou, S., Tannenbaum, R. J., Bridgeman, B., & Cho, Y. (2015). The association between TOEFL iBT® Test Scores and the Common European Framework of Reference (CEFR) levels. *ETS Research Memorandum Series, August*, i–23. <http://www.ets.org/Media/Research/pdf/RM-15-06.pdf> http://www.ets.org/research/policy_research_reports/publications/report/2015/jvgf
- Riazi, A. M. (2016). *The Routledge encyclopedia of research methods in applied linguistics*. Routledge. <https://doi.org/10.4324/9781315656762>
- Schmeiser, C. B. (2009). *Impact of literacy on college and career readiness and STEM success*. <https://all4ed.org/wp-content/uploads/2013/09/ACTPPT041009.pdf>

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

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.
http://samresources.scholastic.com/ResourceManager/previewresource.spr?_page=0&objectId=3218

Stenner, A. J., & Burdick, D. S. (1997). *The objective measurement of reading comprehension: In response to technical questions raised by the California Department of Education Technical Study Group*. files/1012/Stenner und Burdick - 1997 - The objective measurement of reading comprehension.pdf

Stenner, A. J., Sanford-Moore, E. E., & Burdick, D. (2007). *The Lexile Framework® for Reading: Technical report* (Issue April). https://metametricsinc.com/wp-content/uploads/2017/07/Stenner_Burdick_Sanford_Burdick-The_LFR_Technical_Report.pdf

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*. https://metametricsinc.com/wp-content/uploads/2017/07/Stenner_Does_the_Reader_Comprehend_the_Text....pdf

Swartz, C. W., Hanlon, S. T., & Stenner, A. J. (2011). *Text complexity of English international newspapers in a world preparing for college and career*.

Uys, M., Walt, J. Van Der, Berg, R. Van Den, & Botha, S. (2007). *English medium of instruction: A situation analysis*. 27(1), 69–82. <https://files.eric.ed.gov/fulltext/EJ1150096.pdf>

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*, 7(6), 767–770. <https://doi.org/10.1021/ed077p767>

VERBI software. (2019). *MAXQDA 2020 [Computer software]*. VERBI software. <http://maxqda.com>

Weir, C., Yang, H., & Yan, J. (2000). Establishing the parameters of EAP reading. In *An empirical investigation of the componentiality of L2 reading in English for academic purposes* (pp. 14–51). Cambridge University Press.

Wilkins, C., Hartman, J., Howland, N., & Sharma, N. (2010). How prepared are students for college level reading? Applying a Lexile®-based approach. In *Issues & Answers Report*. <http://ies.ed.gov/ncee/edlabs>

Williamson, G. L. (2004). 1350L Lexile: Matching readers to texts: Student readiness for postsecondary options. *The Lexile Framework for Reading*, 1–30. https://metametricsinc.com/wp-content/uploads/2017/07/StudentReadinessforPostsecondaryOptionsv4_1_MetaMetricsWhitepaper.pdf

Williamson, G. L., Koons, H., Sandvik, T., & Sanford-Moore, E. E. (2012). *The text complexity continuum in grades 1–12*. https://metametricsinc.com/wp-content/uploads/2017/09/Research_Brief-Text_Continuum_Grades_1-12_1.pdf

Williamson, G. L., Koons, H., Sandvik, T., Sanford-Moore, E. E., Koons, H., Sandvik, T., & Sanford-Moore, E. E. (2012). *The text complexity continuum in grades 1–12*. https://metametricsinc.com/wp-content/uploads/2017/09/Research_Brief-Text_Continuum_Grades_1-12_1.pdf

Williamson, G. L., Sandvik, T., Stenner, A. J., & Johnson, A. (2016). *Complexity of university texts in the United Kingdom*. 0–3. https://metametricsinc.com/wp-content/uploads/2018/01/MM_UK_University_Research_Digital.pdf

Wright, B. D., & Stenner, A. J. (1999). One fish, two fish: Rasch measures reading best. *Popular Measurement, Spring*, 34–38. [files/1027/Wright und Stenner - 1999 - One fish, two fish Rasch measures reading best.pdf](#)

APPENDIX A: STUDENT QUESTIONNAIRE

Questionnaire – Research into reading project

You have just completed the Aptis Advanced Reading Test. This questionnaire is designed to gather some information about test-takers.

The questionnaire should not take longer than 5 minutes to answer.

The results of the questionnaire will only be used for research purposes and all answers will be treated anonymously and confidentially.

Should you have any questions about this study, please contact [REDACTED].

Thank you for your participation!

***Required**

1. What is your candidate reference number? *

2. Gender *

Mark only one oval.

☐ female

☐ male

☐ non-binary

☐ prefer not to say

☐ Other: _____

3. Age *

Mark only one oval.

- ☐ 15-20
☐ 21-25
☐ 26-30
☐ 31-35
☐ older than 35
☐ prefer not to say

4. What is your first language? (more than one answer possible) *

Tick all that apply.

- ☐ Afrikaans
☐ Arabic
☐ Croatian
☐ English
☐ Estonian
☐ French
☐ Latvian
☐ Lithuanian
☐ German
☐ Greek
☐ Hebrew
☐ Hungarian
☐ Italian
☐ Polish
☐ Russian
☐ Serbian
☐ Slovenian
☐ Spanish
☐ Turkish

Other: ☐ _____

5. What is your intended major? *

Mark only one oval.

- ☐ International Business Administration
- ☐ Contemporary Communications
- ☐ English Language and Literature
- ☐ International Relations and Development
- ☐ Psychology
- ☐ Theology
- ☐ Engineering
- ☐ Other: _____

6. What is your year of study for your intended major? *

Mark only one oval.

- ☐ year 1
- ☐ year 2
- ☐ year 3

7. In general, how difficult do you find the reading materials in the courses for your intended major? *

Mark only one oval.

- ☐ 1 very easy *Skip to question 8*
- ☐ 2 rather easy *Skip to question 8*
- ☐ 3 rather difficult *Skip to question 9*
- ☐ 4 very difficult *Skip to question 9*

8. Why do you find the reading materials in the courses for your intended major easy? (more than one answer possible) *

Tick all that apply.

- ☐ familiar words and phrases
☐ familiar grammatical structures
☐ familiar topics and concepts
☐ length of texts
☐ structure and organisation of texts

Other: ☐ _____

Skip to question 12

9. Why do you find the reading materials in the courses for your intended major difficult? (more than one answer possible) *

Tick all that apply.

- ☐ unfamiliar words and phrases
☐ unfamiliar grammatical structures
☐ unfamiliar topics and concepts
☐ length of texts
☐ structure and organisation of texts

Other: ☐ _____

10. How do you cope with difficult reading materials in the courses for your intended major (more than one answer possible)? *

Tick all that apply.

- ☐ ask course instructor
☐ ask student colleagues
☐ use a dictionary
☐ consult other online resources (e.g. Google, Wikipedia etc.)
☐ guess meaning from context
☐ translate into first language
☐ look up resources in first language
☐ take an English course

Other: ☐ _____

11. What do your course instructors do to help you with the texts?

12. How similar were the texts in this reading test to the texts you need to understand in the courses for your intended major? *

Mark only one oval.

- ☐ 1 very similar
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5 very different
- ☐ Other: _____

13. Please choose which of the following statements about the ability to read texts in English apply to you. (more than one answer possible) *

Tick all that apply.

- ☐ I can understand texts that consist mainly of high frequency everyday or job-related language. I can understand the description of events, feelings and wishes in personal letters.
- ☐ I can read articles and reports concerned with contemporary problems in which the writers adopt particular stances or viewpoints. I can understand contemporary literary prose.
- ☐ I can understand long and complex factual and literary texts, appreciating distinctions of style. I can understand specialised articles and longer technical instructions, even when they do not relate to my field.
- ☐ I can read with ease virtually all forms of the written language, including abstract, structurally or linguistically complex texts such as manuals, specialised articles and literary works.

APPENDIX B: TEST ADMINISTRATION

GUIDELINES

Administering the APTIS Reading Test

Preparation:

1. Print out the consent forms.
2. Print out the keycode slips.
3. Switch on all computers that will be used for the test, plus some backups if available.
4. Put a consent form on each table.
5. Open the internet browser on all computers (if possible use Google Chrome).
6. Enter the website: <https://aptis-replay.bc.janisoncloud.com> (make sure it is https rather than http).
7. If possible, write the following link on a whiteboard or blackboard or project it onto a screen:
http://bit.ly/Reading_Study_2019

Administration:

1. When the students arrive, tell them to sit down in front of one of the computers.
2. Greet them and thank them for their participation.
3. Ask them to turn off their mobile phones.
4. Ensure that they do not have any books etc. on the table. Students only need a pen.
5. Tell them that they will complete a grammar/vocabulary and a reading test and fill in a questionnaire.
6. Ask them to read and sign the consent form.
7. Tell them that the reading test consists of four parts with up to ten items each.
8. Tell them that random participants were created to ensure anonymity. They should not worry whether the information on screen about their name and birthday is accurate. They should just confirm whenever asked that the name and date on screen are correct.
9. Tell them it is crucial that they take the test seriously.
10. Tell them that you are only there to help whenever any technical problems arise and not to help with any questions.
11. Students will receive their results some weeks after the test (detailed scores and CEFR level). They need to note down their candidate reference number to receive their results.
12. Hand out the keycode slips.
13. Tell them to click "Start a Test".
14. Tell them that they should enter the session code.
15. Tell them that they should enter the student code for the reading component.
16. Remind students to read the instructions carefully and to answer all questions. Sometimes the questions do not fit on one screen and the students need to scroll down to see all questions.
17. Tell them to raise their hand once they are finished with the reading test.
18. Once a student has finished the reading test, make sure they have actually finished and submitted it.
19. Open the website [REDACTED] or tell students to enter the website and ask them to fill in the questionnaire.
20. Once a student filled in the questionnaire and noted down their candidate reference number, they are free to go.
21. Restart the website if other participants are coming in. If not, turn off the computer.

What if...

1. ...the website does not load:
 - a. Try a different browser.
 - b. Restart the computer.
 - c. Use a different computer.
2. ...the website crashes during the test:
 - a. Reload the website.
 - b. Use a spare keycode (this will start a new test).
3. ...the keycode does not work:
 - a. Check whether the code has been entered correctly.
 - b. Use a spare keycode.
4. ...any other problem occurs: email us at [REDACTED] or call us on [REDACTED]

APPENDIX C: INFORMATION DOCUMENT FOR READING TEST PARTICIPANTS



Date: [REDACTED]

INFORMATION SHEET

We have approached you because we, the LTRGI (Language Testing Research Group Innsbruck), are conducting a research project to investigate how well students in English as a medium of instruction (EMI) courses are able and ready to deal with the course materials that they are encountering when arriving at university. We would be very grateful if you would agree to take part.

You will be asked to complete an English reading test and fill in a questionnaire. You are free to withdraw from the study at any time. At every stage, your name will remain confidential. The data will be kept securely and will be used for academic purposes only.

If you have any questions about the study, please feel free to contact the research project leader, [REDACTED], who can be contacted at [REDACTED] or by phone on [REDACTED].

[REDACTED]

University of Innsbruck
Institute of Subject-Specific Didactics
Language Testing Research Group Innsbruck (LTRGI)
www.uibk.ac.at/dis/ltrgi
Innrain 52d, Room no. 40508, 6020 Innsbruck, Austria

APPENDIX D: INFORMATION DOCUMENT FOR INTERVIEW PARTICIPANTS

Date: [REDACTED]

INFORMATION SHEET

Text and task features of university reading materials in relation to Aptis and Lexile scores

Dear participant,

We, the *Language Testing Research Group Innsbruck* (LTRGI), have approached you because we are conducting a research project into text and task features of university reading materials. In particular, our aim is to find out to what extent language entrance requirements of university programs with English as a medium of instruction (EMI) correspond to the actual demands of university reading materials. In a previous step, we have received reading materials from your institution, and we have conducted a series of lexical analyses to determine the materials' level of difficulty. In a second step, we now set out to examine the use of reading materials in university classes from the instructors' perspective. We are most grateful that you agreed to take part in our study.

You are asked to partake in an interview and answer a number of questions on your choice and use of reading materials, and how you support students with the reading materials used in your course. We will ask you to answer all questions honestly and to the best of your knowledge. To analyze the data, it will be necessary to audio- and video-record the interviews. The data will be stored securely and will be used for academic purposes only. To avoid health risks, we will conduct all interviews online.

You are free to withdraw from the study at any time. At every stage, your name and other personal information will be anonymized and treated confidentially. If you have any questions about the study, please feel free to contact the research project leader, [REDACTED], at

[REDACTED] or by phone on [REDACTED].

[REDACTED]

University of Innsbruck
Institute of Subject-Specific Didactics
Language Testing Research Group Innsbruck (LTRGI)
www.uibk.ac.at/dis/ltrgi
Innrain 52d, Room no. 40508, 6020 Innsbruck, Austria

APPENDIX E: CONSENT FORM FOR STUDY PARTICIPANTS



Department of Language Didactics/Language Testing Research Group

Consent Form

Project title: **Text features of EMI university course materials in relation to Aptis and Lexile reading scores**

1. I have read and had explained to me the Information Sheet relating to this project.
2. I have had explained to me the purposes of the project and what will be required of me, and any questions have been answered to my satisfaction. I agree to the arrangements described in the Information Sheet in so far as they relate to my participation.
3. I understand that my participation is entirely voluntary and that I have the right to withdraw from the project any time.
4. I have received a copy of this Consent Form and of the accompanying Information Sheet.

Name:

Signed:

Date:

APPENDIX F: BIODATA QUESTIONNAIRE AND INTERVIEW GUIDELINES

Note: The guidelines for the institution in Egypt are shown. The guidelines for the other institutions were the same except for section D, which included the specific results for each institution.

Interview guideline: Egypt

1. Biographic data

Initials (First name/ Last name) ____/____

Interviewee number: I ____

Gender:

Female	<input type="checkbox"/>
Male	<input type="checkbox"/>
Other	<input type="checkbox"/>
Prefer not to say	<input type="checkbox"/>

Recruited from a Higher Education institution in:

Austria	<input type="checkbox"/>
Lithuania	<input type="checkbox"/>
Egypt	<input type="checkbox"/>

Field of study: _____

2. Inclusion criteria

All inclusion criteria need to apply; otherwise, the participant needs to be excluded from the study.

	Yes	No
18 years or older	<input type="checkbox"/>	<input type="checkbox"/>
Experience as an instructor in EMI programs	<input type="checkbox"/>	<input type="checkbox"/>
Sufficient knowledge of English or German for conducting the interview	<input type="checkbox"/>	<input type="checkbox"/>
Signed consent form	<input type="checkbox"/>	<input type="checkbox"/>
Health and safety measures can be respected	<input type="checkbox"/>	<input type="checkbox"/>
Satisfactory technical quality of the online interview	<input type="checkbox"/>	<input type="checkbox"/>
Satisfactory audio/video recordings of the interview	<input type="checkbox"/>	<input type="checkbox"/>
Candidate needs to be excluded from the study	<input type="checkbox"/>	<input type="checkbox"/>

Date (DD/MM/YYYY) ____/____/____

Researcher _____

3. Variables: Background information

1. How old are you? (age on the day of the interview)
2. How many years of teaching experience in higher education do you have?
3. Which subject area do you teach?
4. At which institution / department do you teach?
5. How many years of experience do you have in teaching EMI classes and programs?

4. Semi-structured interviews: English version

Guidelines for the semi-structured interviews

[Introduction]

Thank you very much for participating in our study as an interview partner. We, the Language Testing Research Group Innsbruck, are conducting a research project into text and task features of university reading materials. Our aim is to find out to what extent language entrance requirements of university programs with English as a medium of instruction correspond to the actual demands of university reading materials. The participants from your university took a reading test to determine their level of reading comprehension. We have also received “real life”-reading materials from Kairo, and we have conducted a series of lexical analyses to determine the complexity of the texts based on the so-called Lexile score. We have then compared the demands of the reading assignments at your institution and the students’ actual performance in reading.

Now, we would like to find out more about the use of course reading resources in university classes from the lecturer’s perspective. We would like to know how and why instructors choose reading materials for their classes and how they work with it. So, I will ask you a number of questions on how you use the reading materials in your university classes. We kindly ask you to answer all questions honestly and to the best of your knowledge. If you wish to take a break or if you would like to stop the interview, you can do so at any point. Also, if there is a question you would rather not answer, just let us know. Do you understand the procedure and our expectations? Are you ready to start? Are there any remaining questions?

[Transition a]

The first couple of questions are about how you select reading materials and the role they play in your teaching. By “reading materials” we mean all of the written texts you use in your teaching. These may include individual texts in textbooks, supplementary readings, online texts, or other reading materials.

(a) How do you select reading texts for your courses?

Guideline to (a): Possible further questions

Q1: What kind of reading materials do you usually use in your teaching and where do you find them (e.g. textbooks, supplementary texts, online texts, library purchases)?

Q2: Do you have a say as to which texts you would like to work with in your classes?

Possibly: Are there any institutional restrictions? Do you rely on what has been used before?

Q3: For which purposes do you use reading materials in your classes (e.g., basic/ further reading, basis for test content, basic reading for group work / presentations, etc.)?

Q4: Do you consider aspects about a text’s context when you select reading materials for your classes? What aspects would that be?

Q5: Do you consider aspects about a text’s language when you select reading materials for your classes? What aspects would that be?

Q6: Do you adapt the reading materials in your courses to the students’ language competence level? *possibly:* To what extent? How?

Q7: Do you take student feedback into account when selecting reading materials?

Q8: Compared to the overall time students need to spend on your course, how much of it do they spend on reading?

Q9: Do you explicitly test reading skills in the course assessment and are reading skills necessary for course completion?

[Transition b]

The following few questions will be about what kind of reading tasks students typically need to complete based on the reading materials. By "tasks", we mean the instructions on what students need to do with the text.

(b) How do you select reading tasks for your course?

Q1: What are typical reading tasks you assign in your courses?

Q2: If you assign a reading task, how do you let your students know what to do with the reading? In writing, orally, or implicitly?

Q3: How do students usually respond to the reading task? In writing, orally, or implicitly?

Q4: How do students usually react to the reading tasks?

[Transition c]

We will now focus on how instructors can help students with understanding challenging reading materials. We will also talk about what kind of institutional support EMI universities might offer students to deal with the linguistic demands of their courses to enhance learning.

(c) What kind of support do you provide if students struggle with text comprehension?

Guideline to (c): Possible further questions

Q1: What do you usually do if you realize that your students struggle to understand the reading materials?

Q2: You have already mentioned that... Do you use other ways of scaffolding to foster reading comprehension?

Q3: In your view, is your university doing enough to support the language needs of the students?

Q4: What kind of support could EMI universities offer their students to help them deal with the linguistic demands of their studies?

[Transition d]

We will now move on to some questions about how difficult reading tasks are for the students at your university. We have given you a selection of texts from your institution. Now, we will ask you to judge the level of complexity of these texts, assuming that there have been no pre-tasks and the topics have not been covered in class. We will give you between five and ten minutes to skim through the texts. You don't need to read the texts as a whole, but just to have a look at them and try to get a feeling for their level of complexity. While reading, you can switch off your microphone and the camera. Take the time you need and let us know when you're finished. Please rank the texts from the most to the least complex text.

(d) How complex would you estimate the course reading materials to be for the students at your institution?

- Q1: Could you please tell us your ranking from the most to the least complex text?
 Q2: Could you please explain why you estimated these texts to be more or less complex?
 Q3: Which aspects about a text's context did you consider when ranking the reading materials? *Possibly*: Why do you think a text might be challenging for students?
 Q4: Which aspects about a text's language did you consider when ranking the reading materials? *Possibly*: What do you mean by "language" / "difficult" etc.?
 Q5: Which other aspects did you take into account?

Key:

Text	Lexile
01_08_DescartesDiscourse	1660
01_33_The Practice of Science_Process of Science_Visionlearning	1440
01_36_What Are Proxy Data	1210
01_19_How to Make a Line Graph_8 Steps	1000
01_34_The-mean-median-and-mode	790

[Transition e]

We will also discuss what experiences you have made with text complexity as a teacher and to what extent our analyses match your experience.

(e) What is your experience with text difficulty and using reading materials in teaching?

Guideline to (e): Egypt

To start with, I will briefly tell you the results of our analyses. To recap, we tested the students reading skills based on the Lexile score and compared them to the text complexity in "real life" university reading materials, also based on the same score. According to our analyses, instructors at your institution used reading materials that were generally above the students' average language proficiency level. So, the reading materials did not quite match the students' reading skills. Although the Egyptian participant group contained the student with the highest test score of all test takers, we also discovered important differences between individual texts and individual test takers. Of all the reading materials we collected and analysed from your institution, about two thirds of the texts were above the level of proficiency of about 50% of the students. This means that half of the students might struggle to understand about two thirds of the texts because they have a lower level of reading skills.

- Q1: Would you have expected that result?
 Q2: How do our results compare to your personal experience as a lecturer?
 Q3: Have you already experienced students expressing concerns about reading materials being (too) difficult / easy/ complex / incomprehensible / technical, etc.?

[Transition f]

With the last couple of questions, we will try to find out what you think instructors and institutions might do differently in the future and what further means of support they might provide.

Based on this study, do you think you will consider different aspects in your future choice of reading materials?

Guideline to (f): Possibly repeat the results from (e)

Q1: *If yes:* What would you do differently? / *If no:* Why not?

Q2: Do you think that our results might affect your expectations towards your students?

Q3: Will you consider implementing further strategies to support student reading in the future?

Q4: Do you foresee these results to have any other effect on your teaching?

Q5: Could you imagine that the results of our study might be discussed in your department or institution?

[Goodbye]

This is the end of the interview. Thank you very much for participating in our study! We are very grateful for your help and the insights you have provided. If you have any further questions about our study, feel free to ask them now or get in touch with us later on.

Research into Reading Grants

COMPARING EMI UNIVERSITY READING MATERIALS WITH STUDENTS' READING PROFICIENCY UTILIZING LEXILE® MEASURES, APTIS TEST RESULTS, STUDENT QUESTIONNAIRE RESPONSES, AND INTERVIEWS WITH LECTURERS

RESEARCH IN EFL READING AND READING ASSESSMENT Report 2, 2022

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

BRITISH COUNCIL
www.britishcouncil.org

METAMETRICS
The Lexile Framework for Reading – Lexile
Matching Readers with Texts
www.lexile.com

© British Council and MetaMetrics 2022
ISSN 2755-2543