LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

EMPIRICAL STUDY

Vocabulary and Processing Speed Explain Reading and Writing Disparities Between Linguistic Groups in Higher Education

Justyna Mackiewicz[a] and Danijela Trenkic[b]

[a]University of Birmingham [b]University of Oxford

Abstract: Selective admissions at universities in the United Kingdom aim to ensure a baseline language competence, yet, despite persistent achievement disparities across linguistic backgrounds, systematic comparisons of linguistic skills underpinning academic success remain rare. This study compared English proficiency among three groups of first-year undergraduates: British (n = 60), European (n = 59), and Chinese (n = 58). Two proficiency skills—reading comprehension and text summarization—were assessed alongside 11 theoretically motivated component measures. Although previously reported gaps between British and Chinese students were replicated, European students performed comparably to British students on several measures, reflecting English proficiency variation across L2 groups consistent with sector-level patterns of academic outcomes. Group differences in reading and writing were explained by underlying component skills, particularly vocabulary knowledge and processing efficiency. Findings carry theoretical and policy implications, highlighting the need for targeted

---

CRediT author statement – **Justyna Mackiewicz**: conceptualization; methodology; investigation; formal analysis; writing – original draft preparation; writing – review & editing. **Danijela Trenkic**: conceptualization; methodology; formal analysis; writing – original draft preparation; writing – review & editing.

A one-page Accessible Summary of this article in nontechnical language is freely available in the Supporting Information online and at https://oasis-database.org.

The research was funded by PhD departmental scholarship, the Department of Education, University of York.

Correspondence concerning this article should be addressed to Danijela Trenkic, University of Oxford, 15 Norham Gardens, Oxford, OX2 6PY, United Kingdom, danijela.trenkic@education.ox.ac.uk

The handling editor for this article was Daniel Isbell.

DOI: 10.1111/lang.70031

linguistic support to address persistent group disparities in academic outcomes within increasingly diverse university settings.

Keywords language proficiency; literacy skills; higher education; academic success; group differences

Introduction

One of the most remarkable aspects of human language is our ability to acquire multiple languages, often in more than one modality, and to do so across the entire lifespan. Nonetheless, we typically achieve higher proficiency, and more closely approximate the patterns to which we have been exposed, in our first language (L1) than in second or additional languages. Moreover, our success in acquiring a second language (L2) is shaped by a range of variables, including its similarity to our L1. A central goal of L2 acquisition research is to explain how and why L2 learning trajectories and outcomes differ from those of L1 development, and why they vary across learner populations. Understanding these differences is not only theoretically important; it can also have far-reaching consequences, particularly in educational settings where students from diverse linguistic backgrounds share a language of instruction. The question becomes even more pressing in contexts where documented disparities in academic success intersect with students' linguistic backgrounds, such as within higher education in the United Kingdom, where the present study is situated.

Similar to other Anglophone countries, the United Kingdom operates a selective higher education system and attracts large numbers of international students. Such systems offer valuable contexts to compare L2 profiles of advanced learners from diverse linguistic backgrounds. Against this backdrop, this study investigated whether, even among students selected for English proficiency, substantial differences persist between L1 and L2 speakers of English, and among L2 speakers of English with different L1s. If so, on which dimensions of proficiency are these differences most evident?

We were particularly interested in text-level reading and writing, as both are complex literacy skills central to academic study. These skills are widely recognised as multicomponent processes that rely on various aspects of linguistic knowledge (lexical, grammatical, phonological and orthographic); fluency in decoding and encoding written language; background knowledge of topics, genres, socio-cultural norms, and academic discourse conventions; and the working memory needed to integrate these elements. This complexity means that even proficient L1 and L2 speakers of English may face challenges in academic reading and writing that do not necessarily reflect linguistic weakness.

Therefore, the second goal of our study was to examine whether disparities in text-level literacy skills among university students in the United Kingdom could be explained by underlying linguistic components (e.g., vocabulary, grammar, etc.), and, if so, which components were most strongly associated with these disparities. To investigate this, we utilized a comprehensive set of theoretically motivated measures that captured lexical, grammatical, phonological, and orthographic knowledge, while also controlling for cognitive and socio-economic variables that may influence performance.

The study addressed these questions by comparing three linguistically distinct groups of first-year undergraduates in the United Kingdom: domestic students who spoke English as their L1; international students who spoke European languages as more linguistically similar to English; and international students who spoke a variety of Chinese, which is more linguistically distant from English. By examining whether meaningful group differences in English language and literacy skills persisted where language proficiency is nominally controlled, and by analyzing the linguistic and cognitive variables underlying skilled reading and writing, this study aimed to support a more nuanced understanding of language proficiency and multilingual students' needs in higher education. We also discuss both theoretical and policy implications, offering evidence-based recommendations to help level the playing field for L2 students.

Higher Education International Students in the United Kingdom: Academic Outcomes

International students constitute a significant proportion of the student population at universities in the United Kingdom. According to the Higher Education Statistics Agency (HESA, 2025b), 732,285 non-domestic students were enrolled during the 2023–2024 academic year, accounting for 25.2% of all students. Although no language background information was collected, nationality data indicate that a large majority of these international students come from countries where languages other than English are predominantly spoken. Thus, these students are likely to use English as a L2.

While many international students in the United Kingdom excel academically, they nonetheless, on average, achieve lower degree outcomes than their domestic peers. This disparity is captured in sector-level monitoring of degree-awarding gaps, defined as differences in the proportion of 'good' (first- or upper-second-class) degrees awarded to different student groups. These degrees are generally regarded as indicators of strong academic performance and are often prerequisites for postgraduate study or graduate-level employment.

Annual reports have consistently shown that students from outside Europe—though not those from within Europe—are less likely to obtain good degrees than their British peers (Borrett & Foster, 2023; Morrison et al., 2005), with Chinese students often faring less well than other international groups (Crawford & Wang, 2015; Iannelli & Huang, 2014). Our analysis of cumulative data from 2014–2024 (HESA, 2025a) confirmed that 32% of students from outside Europe failed to obtain good degrees, compared with 22% of British and 18% of European students.

However, the absence of HESA data on language background and English proficiency means the role of language in these outcome differences remains speculative. A key contention is that, since international L2 students must demonstrate readiness to study in English, they should, in principle, be linguistically prepared for study.

International Higher Education Students in the United Kingdom: English Language Proficiency

Although international L2 students must demonstrate readiness to study in English, a 2020–2021 review of university entry requirements in the United Kingdom found substantial discrepancies between recommended standards of proficiency and those accepted for admission (Pearson, 2020). Using the International English Language Testing System (IELTS) as a yardstick (because it is a government-approved test that provides both overall scores and subscores for listening, speaking, reading, and writing on a 1–9 scale), the study found that only one institution required the recommended score of 7.5 for unconditional entry.[1] This is equivalent to a Common European Framework of Reference for Language (CEFR) level of C1. Instead, most institutions set the overall cutoff at 6.0, with sub-scores at 5.5. This is the government's threshold for degree-level student visas (CEFR B2) in the United Kingdom. The same study found that applicants below the cutoff were usually offered pre-sessional pathways that accepted sub-scores as low as 4.0–4.5 (CEFR B1, bordering on A2). Passing pre-sessionals (as most did) granted students degree-study entry without requiring a secure, independent test to reverify proficiency. Similar to L1 background, data on English proficiency are not centrally collected, so the prevalence of students arriving at below-recommended levels—and whether this varies by language background—remains unknown.

What is clear from available research, however, is that the English proficiency of some students at enrollment presents a barrier to achievement. Trenkic and Warmington's (2019) study found that students arriving with the IELTS-recommended score of 7.5 perform, on average, a whole degree

classification better than their equally capable peers from the same educational and cultural background who arrived with a widely accepted unconditional entry score of 6.5. Furthermore, students entering via pre-sessional pathways underperformed compared to those who met the minimum unconditional scores (Thorpe et al., 2017). Comparable trends are observed in Australia (Eddey & Baumann, 2011; Oliver et al., 2012), additionally documenting that longer pathways are linked to poorer degree–study outcomes, presumably because they aim, but fail, to bridge larger proficiency gaps.

When faced with linguistic demands exceeding their proficiency, students report turning to strategies such as using online translation tools or AI summarization to cope (Vicary & Treffers-Daller, 2024). The need for additional linguistic support is widely recognized by universities. It is, however, rarely tailored to specific needs. Indeed, the provision is often offered universally, to both L1 and L2 English speakers, under the premise that academic English is nobody's first language (Hyland, 2016), covering study skills, academic English, and disciplinary writing norms. This approach rightly acknowledges that higher education literacy demands stretch most students. Yet the assumption that international students (regardless of proficiency) face challenges not dissimilar to those of British students transitioning from A-levels may underestimate the type and magnitude of the difficulties some students encounter. To better gauge needed support, and whether addressing individual differences in English proficiency could reduce group disparities, research is needed to understand incoming students' English skills in relation to theoretical models of language proficiency and skilled reading and writing.

Group Differences in English Language and Literacy Skills

Extensive research has demonstrated the central role of language and literacy skills in academic success at all educational levels, including higher education (Daller et al., 2021; Spit et al., 2025). Yet, despite persistent achievement disparities linked to linguistic background, few studies have systematically compared these skills across student groups in English-medium universities. One of the first studies based in the United Kingdom to do so, and the foundation for the present work, was done by Trenkic and Warmington (2019). Drawing on theoretically informed models of literacy, that study compared the English language and literacy skills of British students and international Chinese students who had met the English proficiency requirements typically accepted for unconditional entry (IELTS 6.5–7.5). Despite this, substantial group differences were observed across a wide range of measures, including academic English skills (e.g., reading comprehension and written summarization) and

components skills (e.g., vocabulary, phonological awareness, phonological retrieval, word-reading accuracy, spelling, and text-reading speed).

Crucially, and consistent with findings from other studies of L2 students in higher education (Daller & Phelan, 2013; Harrington & Roche, 2014; Hu & Trenkic, 2021; Masrai & Milton, 2021; Roche & Harrington, 2013), individual differences in linguistic skills accounted for more than half the variance in academic grades for the Chinese group, whereas the same model did not predict British students' grades. The authors suggested these results support the view that language proficiency constrains academic performance only below a certain threshold and that the English skills of many international students may fall short of this threshold. This limits their ability to realize their academic potential.

Trenkic and Warmington's (2019) study also highlighted the importance of students' incoming language proficiency. That is, regardless of any language learning that happened during their studies, it was Chinese students' initial skills that were critical for their academic success. This aligns with research indicating that, across pre-sessional programs and degree-level study, international students generally improve at a similar rate regardless of starting proficiency, so gaps persist as the linguistic demands of study increase (Schmidtke et al., 2023, 2025). Some studies report narrowing of these gaps, but these typically include students from immigrant backgrounds or long-term residents rather than solely international students (McCarron & Kuperman, 2022).

Large group differences in English language and literacy skills between L1 and L2 speakers are by no means inevitable, as most recently documented by multisite studies comparing advanced university-level L2 learners and native English speakers across countries (Kuperman et al., 2025; Siegelman et al., 2024). However, the populations in these studies (primarily students studying in their first language and within their home countries) limit their applicability to English-medium instruction contexts, where diverse linguistic groups study side by side and where financial pressures may incentivise institutions to adopt flexible language entry requirements to sustain international student recruitment. To what extent meaningful group differences exist is thus an empirical and context-dependent question.

Within higher education in the United Kingdom, it remains unclear whether the large linguistic gaps observed for Chinese students, whose first language is typologically distant from English (Trenkic & Warmington, 2019), generalize to other international groups, particularly those from linguistically closer backgrounds. Linguistic distance has already been shown to influence adult language learning (Schepens et al., 2020) and literacy processes such as

reading fluency (Kuperman, 2025; Nisbet et al., 2022). Therefore, although universities in the United Kingdom aim to ensure a common baseline of English proficiency, discrepancies between recommended and accepted entry standards mean that, on average, students from more distant linguistic backgrounds could be entering with lower levels of linguistic readiness than peers whose first languages are closer to English. To investigate this, the present study focuses on European students, whose L1s are typologically closer to English than is Chinese and, importantly, whose academic outcomes are comparable to their British peers. Prior research suggests that students with European L1s tend to have a larger English vocabulary than Chinese students (Larson, 2017) and report fewer language-related difficulties (Senyshyn et al., 2000; Trice, 2003). However, no study has systematically compared Chinese, European, and British students within the same educational context across a comprehensive range of linguistic skills known to underpin academic performance. The present study addresses this gap.

Conceptual Framework and the Present Study

Fundamentally, education relies on language and literacy skills: In particular, weak reading comprehension limits content learning and a limited ability to express oneself in writing undermines assessment performance (August & Shanahan, 2006). These text-level literacy skills draw on multiple components, including core linguistic proficiencies (e.g., vocabulary and grammar); lower-level literacy skills (e.g., decoding and spelling); and background knowledge spanning topic familiarity, socio-cultural understanding, genre conventions, and disciplinary writing norms. Theoretical models of skilled reading and writing emphasize their multicomponential nature and the intricate interaction of these elements coordinated within the limits of working memory (Hayes, 2012; Perfetti & Stafura, 2014; Schoonen et al., 2009). This complexity means that even proficient L1 and L2 speakers of English may encounter challenges with academic literacy. For individuals with limited language proficiency, however, these challenges are likely to be compounded.

Although reading and writing involve processes beyond language, their foundation rests on lexical, grammatical, phonological, and orthographic knowledge. In reading comprehension, readers must identify words, access meanings, and parse syntax to build a coherent mental model. In writing, authors must transform thoughts into structured language and encode it in written form. The fluency with which these processes are executed—often referred to as processing efficiency—is equally critical. When linguistic representations are weak or slow to access, cognitive effort is diverted from higher-order tasks,

such as integrating ideas across a text or planning coherent discourse structure, impairing both reading comprehension and text writing.

Lexical and grammatical knowledge tend to be stronger predictors of reading and writing performance in a L2 than in a L1, reflecting lower average proficiency and greater variability among L2 speakers of a language (Jeon & Yamashita, 2014; Kojima et al., 2022). Consequently, difficulties with reading and writing in a L2 often stem from limits in linguistic proficiency rather than from the literacy skills themselves. Nevertheless, because successful reading and writing draw on broader cognitive and strategic resources, advanced L2 users can often compensate for remaining linguistic gaps to achieve levels of skill comparable to L1 speakers on some aspects of literacy performance, such as reading comprehension (e.g., Siegelman et al., 2024).

It remains an open question whether, in universities where diverse linguistic groups study side by side, differences in linguistic component skills lead to meaningful disparities in academic literacy. Students are typically selected for linguistic readiness, either through standardized tests or prior academic achievement in English-medium education, suggesting that groups might converge in academic literacy even when differences remain on component skills. Alternatively, typically accepted proficiency levels may leave large residual gaps that could affect performance on complex reading and writing tasks, particularly under time pressure, with students from more linguistically distant backgrounds potentially being more affected.

This study explored these issues by comparing three groups of first-year undergraduates in the United Kingdom: British, European, and Chinese. The investigation focused on the complex literacy skills of reading comprehension and written summarization and their supporting components (i.e., vocabulary, grammar, word decoding, spelling, phonological awareness and retrieval) as well as text-reading and writing efficiency. Parental education, nonverbal intelligence, and working memory were controlled for. Building on the work of Trenkic and Warmington (2019), who found large gaps in all aspects of English proficiency between British and Chinese students, the present study tested whether such large disparities can be replicated and whether they extend to European students studying in the same context. We further examined whether any observed group differences in reading and writing would persist once core aspects of proficiency were accounted for.

The study addressed two main questions:

RQ1 . In which aspects of English proficiency do British, European, and Chinese first-year students at universities in the United Kingdom differ, and to what extent?

RQ2. After accounting for individual differences in linguistic and cognitive components, do group differences in reading and writing persist, and which components are most strongly associated with performance?

Method

Participants

All participants (N = 177; 123 female, 54 male) were first-year undergraduates at a research-intensive university with English language entry requirements typical of the sector in the United Kingdom. For direct entry, students had to have an overall IELTS score of at least 6.0, with no subscore below 5.5 (CEFR B2). For entry via pre-sessional pathways, students had to have an overall IELTS score of at least 5.0, with no subscore below 4.5 (CEFR B1).

The sample comprised of 59 domestic students who identified English as their first language (i.e., British group) and 118 international students who did not speak English as a L1. Among these, 58 reported a variety of Chinese as their L1 (i.e., Chinese group) and 60 spoke a European language as their L1 (i.e., European group). The Chinese group included 40 Mandarin speakers, 12 Cantonese speakers, and six who reported their first language as Chinese, without further specification. In the European group, there were 33 speakers of Romance languages, including Italian (11), Spanish (9), French (9) and Romanian (4). There were nine speakers of Germanic languages, including Norwegian (5), German (3) and Danish (1), and 11 speakers of Balto-Slavic languages including Lithuanian (4), Polish (3), Russian (2), Slovak (1) and Czech (1). In addition, the sample included five Greek speakers, one Hungarian speaker, and one Finnish speaker. The mean age was 18.67 years (SD = 0.92). None of the participants reported having a diagnosed reading or learning disability. Participants were recruited through posters advertising the study and received payment for their participation.

Design and Procedures

Participants completed a battery of cognitive, language, and literacy tests within the first semester of enrollment. A short questionnaire collected demographic and language background data. All measures were administered

individually except the vocabulary size test, which was conducted in groups of up to 20 students under exam conditions. Individual sessions took between 50–60 min, and the group vocabulary sessions took 30–50 min. All tests were administered in English. The study was approved by the Ethics Committee of the Department of Education, University of York.

Materials and Measures

12 measures spread across 10 tasks assessed the participants' general cognitive abilities as well as their English vocabulary, grammar, phonological processing, and literacy skills. In addition, participants reported the educational level of their parents. This was used as an index of socio-economic status.

Cognitive Skills and Socio-Economic Status

Nonverbal cognitive abilities and socio-economic status can affect individual differences in language and literacy skills, the key dependent variables in this study. Therefore, fluid intelligence, working memory, and parental education were included as background predictors in the analyses.

Fluid intelligence. Nonverbal reasoning was assessed using the matrix reasoning subtest from the 4th U.K. edition of the Wechsler Adult Intelligence Scale (WAIS–IV U.K.; Wechsler, 2010). While taking this test, participants selected the correct option to complete a visual pattern. The test contained 26 items, with a reported internal reliability coefficient of .87. The sum of correct answers was used in regression analyses.

Working memory. The digit backward span subtest from the WAIS-IV U.K. (Wechsler, 2010) assessed working memory capacity. Participants recalled orally presented digit sequences in the reverse order. The task began with two two-digit sequences. The length of the sequences increased whenever at least one sequence was recalled correctly. The mean length of the last two correctly recalled sequences was used in the analyses. The maximum sequence length was eight digits. The test has a reported split-half internal reliability of .92.

Socio-economic status. Participants reported their parents' highest education on a 5-point scale (1 = *below secondary*; 2 = *secondary*; 3 = *post-secondary*; 4 = *university*; 5 = *postgraduate*). The mean of each student's mother's and father's educational attainment (i.e., parental education) was used in the analyses.

English Vocabulary and Grammar

Lexical and grammatical knowledge are established indicators of overall language proficiency and underpin both oral and written comprehension and pro-

duction. Indeed, previous research identifies them as robust predictors of complex literacy skills (Jeon & Yamashita, 2014; Kojima et al., 2022).

Vocabulary. The online version of the Vocabulary Size Task (Nation & Beglar, 2007) assessed vocabulary knowledge. From four options, participants selected the correct definition for 140 target words presented in minimal sentence contexts. The sum of correct answers was used in analyses. Based on the raw score, the software also estimated the overall vocabulary size (i.e., the number of root word families known) for each participant. The reported internal reliability of the Vocabulary Size Task is .94.

Grammar. Grammatical knowledge was assessed using the test from Hartshorne et al.'s (2018) study, which required participants to make grammaticality judgements across a range of constructions (e.g., passives, clefts, subject-verb agreement, relative clauses, prepositions). The test contained 95 scored (critical) items worth one point each. Reported internal reliability of this tool is .86.

Phonological Processing

Phonological skills are closely linked to reading development in both children and adults. Recent work indicates that the phonological skills upon enrollment of an international student in pre-sessional programs predict their development of degree-level reading skills (Schmidtke & Moro, 2021).

Phonological awareness. The Elision task from the Comprehensive Test of Phonological Processing (CTOPP, Wagner et al., 1999) assessed the participants' ability to remove specific segments from provided words to form new words (e.g., "say *bold* without /b/."). The test contained 20 items, and the sum of correct answers was used in analyses. The reported internal reliability of this task for adults (18+) is .93.

Phonological retrieval. The Rapid Automated Naming of digits (RAN digits) from the York Adult Assessment–Revised battery (YAA-R; Warmington et al., 2013) assessed the participants' speed of phonological retrieval, a key skill for word-reading fluency. To complete this task, participants named 50 digits arranged in 10 rows of five as quickly and accurately as possible. The number of digits named per second (RAN rate) was used in analyses.

Reading and Writing Skills

Reading and writing in the language of instruction are prerequisites for academic success. We measured the participants' word-reading efficiency, text reading rate, reading comprehension, and ability to produce written summaries of previously read text.

Word-reading efficiency. The Sight Word Efficiency (SWE) subtest from the revised Test of Word Reading Efficiency (TOWRE-2, Torgesen et al., 2012) assessed fluency in recognizing high-frequency words. Participants read a list of 104 words as quickly and accurately as possible. The number of words read correctly in 45 s was used in analyses. Reported internal consistency reliability for the SWE is .95.

Reading comprehension. The comprehension component of the Nelson–Denny Reading Test (Brown et al., 1993, Form G) assessed reading comprehension. It included seven passages followed by multiple-choice questions. To simulate the demands typical of higher education, the test was time-limited to 15 min. The maximum raw score was 38, and the reported parallel-form reliability of this tool is .81.

Text-reading rate. Reading rate (i.e., words per minute) of the participants was assessed using a 492-word, nonfictional text from YAA-R (Warmington et al., 2013). Participants read the text silently focusing on comprehension to support a subsequent written summary.

Writing task. After reading the YAA-R text, participants had 10 min to write a summary without referring back to it. Three measures were derived: spelling error rate (i.e., number of errors per 100 words), indexing orthographic knowledge; text length (i.e., number of words), indexing writing efficiency; and summarization quality, measured as the number of accurately summarized content points, indexing global summarization skills.

Hypotheses and Analyses

Given that academic performance is strongly predicted by language proficiency at entry, that European international students tend to achieve stronger academic outcomes than their non-European peers, and that greater linguistic distance between a learner's first language and English increases the time and effort required to achieve comparable levels of proficiency, European students might be expected to arrive with stronger mastery of English than non-European students. At the same time, selective recruitment could minimize differences, making them inconsequential for literacy skills critical to academic performance.

To address RQ 1—whether and on which aspects of English proficiency the British, European and Chinese groups differ—each literacy and component skill score was treated as a dependent variable in separate linear regression models, with student group (British/European/Chinese) serving as a categorical predictor. When homoscedasticity was met, inference relied on the conventional covariance matrix. Pairwise group comparisons were estimated with emmeans (Lenth, 2025) and Bonferroni-adjusted for multiple compar-

isons. When heteroscedasticity was detected, heteroskedasticity-consistent 3 (HC3) robust standard errors were used. Omnibus effects were tested with a robust Wald F, and pairwise comparisons were computed with the HC3 covariance and were also Bonferroni-adjusted.

To address RQ 2—whether group differences in reading and writing persist once component skills are accounted for—each literacy skill score was treated as a dependent variable in separate hierarchical multiple regressions, with predictors entered in blocks. The first block included matrix reasoning, backward digit span, and parental education to control for cognitive and socio-economic variation. The second block added vocabulary and grammar as core language proficiency indicators. Task-specific component skills (e.g., phonological processing for word reading) were then entered to test whether they explained additional variance beyond vocabulary and grammar. Student group was entered last in order to assess whether group effects persisted after accounting for all predictors. We examined incremental contributions (ΔR^2), and, for final models, calculated each variable's strength of association with the outcome when other variables were held constant (partial R^2). We also calculated the unique variance explained by each (part R^2). Descriptive statistics, linear regression outputs, and summaries of incremental contributions are reported in the main text; final models are in the Supporting Information online.

For categorical demographic variables, chi-square tests examined group associations. For continuous variables violating normality and involving two groups, Mann–Whitney (Wilcoxon rank–sum) tests were applied.

Sample size was determined by practical considerations of testing capacity and available funding. Sensitivity analyses (which can be found in Appendix S7 in the Supporting Information online) indicated that, with 177 participants and a significance level of .05, simple linear regressions had 80% power to detect R^2 values of approximately .052 ($f^2 = .055$). Stepwise models had 80% power to detect incremental ΔR^2 of about .045 for single-predictor blocks, .054 for two-predictor blocks, and .060 for three-predictor blocks, corresponding to $f^2 \approx .045$–.060 (i.e., small-to-moderate effect sizes, closer to small). The study was thus well-powered to detect small-to-moderate incremental effects, though very subtle effects (e.g., $\Delta R^2 < .04$) may not have been reliably detected. Pairwise comparisons using Bonferroni-adjusted significance levels ($\alpha = .017$) were sensitive to medium effects ($d \approx 0.60$, two-tailed). Overall, the study was therefore adequately powered to test whether the large effects observed in Trenkic and Warmington's (2019) study were replicable and generalizable to other international student groups.

Transparency and Openness

The data set is available on the Open Science Framework (https://osf.io/rdn6s/). All analyses were conducted in R (R Core Team, 2025).

Results

Participant and Group Characteristics

Table 1 summarizes the demographic characteristics of the participants.

Chi-square test for categorical variables and linear regressions for continuous variables confirmed that the student groups were well matched for age, gender, and parental education. There was no association between gender and student group, $\chi^2(2, n = 177) = 0.50$, $p = .780$, Cramer's V = .05 (negligible effect), nor a significant age difference, $R^2 = .027$, $F(2,174) = 2.41, p = .093, f^2 = 0.03$ (small effect). For parental education, the original five-point scale was collapsed to four levels ("secondary or lower," "postsecondary," "university," and "postgraduate") because of empty cells. There was no significant association between group and education level for either fathers, $\chi^2(6, n = 172) = 11.62$, $p = .071$, Cramer's $V = .18$, or mothers, $\chi^2(6, n = 173) = 7.20$, $p = .303$, Cramer's $V = .14$ (both small effects).

Table 2 summarizes the European and Chinese groups' experience with English. A slightly higher proportion of the European group reported prior education through English-medium instruction (EMI) and residence in an English-speaking country (ESC), but neither difference was significant; for EMI: $\chi^2(1, n = 118) = 0.31$, $p = .578$, Cramer's $V = .05$; for ESC: $\chi^2(1, n = 118) = 0.72, p = .396$, Cramer's $V = .08$. However, the Chinese group reported an earlier onset of English learning (median age six) than the European group (median age seven), and this difference was significant ($U = 2177.50$, $p = .010$, rank biserial $r = .273$, medium effect size).

Despite this later start, the European group reported higher English proficiency. Self-reported qualifications were mapped onto the Common European Framework of Reference (CEFR, 2025). Among 44 students in the European group with declared test scores, 6.8% reported being at level C2 (the highest level), 63.6% at C1, and 29.6% at B2. Among the 57students in the Chinese group, none reported being at C2; 33.3% were at C1, 64.9% at B2, and 1.8% at B1. The association between group and CEFR level was significant, $\chi^2(3, N = 101) = 15.83$, $p < .001$, Cramer's V = 0.40 (medium effect).

Table 1 Participant demographics by group

	N	Gender		Age		Father's education				Mother's education			
Group		M	F	Mean	*SD*	E1	E2	E3	E4	E1	E2	E3	E4
British	59	16	43	18.49	0.94	15	8	25	10	12	11	22	14
European	60	19	41	18.67	0.86	9	11	19	20	10	8	27	13
Chinese	58	19	39	18.86	0.95	9	8	30	8	7	11	32	6
All	177	54	123	18.67	0.92	33	27	74	38	29	30	81	33

Note. M = male, F = female, E1 = secondary education or lower, E2 = postsecondary education, E3 = university degree, E4 = postgraduate degree. Five data points were missing for father's education (1 British, 1 European, 4 Chinese) and four for mother's education (2 European, 2 Chinese).

Table 2 Chinese and European participants' experience with English

		European ($n = 60$)	Chinese ($n = 58$)
EMI experience (count)		31	27
ESC experience (count)		27	17
Length of Residence in ESC (years)	Minimum	0	0
	Maximum	10	6
	Median	0	0
	Mean (SD)	0.84 (3.21)	0.60 (1.57)
Age when English first taught in school (years)	Minimum	3	2
	Maximum	14	14
	Median	7	6
	Mean (*SD*)	7.15 (2.30)	6.28 (2.48)
English language proficiency, CEFR levels (count)	C2	3	0
	C1	28	19
	B2	13	37
	B1	0	1
	No information	16	1

Note. EMI = English-medium instruction; ESC = English-speaking countries; CEFR = Common European Framework of Reference for Languages.

Group Comparisons: Cognitive Skills

No significant association of student group with either matrix reasoning or backward digit span was observed, suggesting that the groups had comparable levels of fluid intelligence and working memory capacity (Table 3).

Group Comparisons: English Vocabulary, Grammar and Phonological Skills

However, student group was strongly associated with the key English proficiency components of vocabulary, grammar, and phonological processing (see Table 3). The British group had substantially larger vocabulary (16.4k word families) than the European (11.4k) and Chinese (7.4k) groups; Pairwise comparisons (Table 4) confirmed large effect sizes. The British group was also stronger on grammar, being slightly over the European group and substantially over the Chinese group. For phonological awareness (elision) and retrieval (RAN), the British group outperformed only the Chinese group, with large and medium effects, respectively. No differences emerged between British and European groups. The European group outperformed the Chinese group on vocabulary, grammar, and phonological awareness, all with large effects.

Table 3 Cognitive abilities, English language knowledge and phonological skills by group

Measure	Group	Mean	*SD*	Min	Max	*F*(2,174)	*p*	R^2	f^2
Fluid intelligence (Matrix reasoning)	British	20.66	3.14	13	25	2.08	.128	.023	0.02
	European	19.98	2.89	10	25				
	Chinese	21.02	2.24	14	25				
Working memory (Digit backward span)	British	4.93	1.31	2.5	7.5	1.52	.224	.015	0.02
	European	4.71	0.99	2.5	7.5				
	Chinese	5.07	1.32	3.0	8.0				
Vocabulary size (raw score)	British	118.75	7.67	104	135	212.27	**<.001**	.728	2.68
	European	103.85	10.89	77	127				
	Chinese	73.76	14.87	52	114				
Vocabulary size (estimate, N of word families)	British	16,415	3,896	10,400	25,700	–	–	–	–
	European	11,393	2,773	7,700	20,700				
	Chinese	7,441	1,685	5,200	13,700				

(*Continued*)

Table 3 (Continued)

Measure	Group	Mean	*SD*	Min	Max	*F*(2,174)	*p*	R^2	f^2
Grammar (raw score)	British	91.31	2.96	83	95	56.87	**<.001**	.478	0.92
	European	89.67	3.86	77	95				
	Chinese	80.93	6.73	54	92				
Phonologial awareness (Elision score)	British	17.56	2.20	12	20	14.85	**<.001**	.146	0.17
	European	17.63	2.34	9	20				
	Chinese	15.50	2.66	8	20				
Phonological retrieval (RAN rate, digits)	British	2.79	0.50	1.72	4.17	4.15	**.017**	.045	0.05
	European	2.59	0.47	1.61	3.85				
	Chinese	2.49	0.73	1.56	5.56				

Note. nBritish = 59, nChinese = 58; nEuropean = 60 (except matrix reasoning and vocabulary = 59). Due to one missing data point, the degrees of freedom for matrix reasoning and vocabulary were (2, 173). RAN = Rapid naming. For vocabulary size and grammar, HC3 heteroskedasticity-consistent standard errors were used for inference due to unequal variances and Wald F statistics is reported. f^2 = Cohen's f^2, values of 0.02, 0.15 and 0.35 indicate small, medium and large effects, respectively (Cohen 1988). Bolded p values are statistically significant ($p < .05$).

Table 4 Group contrasts on English language knowledge and phonological skills

Measure	Comparison	*B*	*SE*	95% CI	*t*	p_{bonf}	*d*
Vocabulary size (raw scores)	British–European	14.90	1.75	[10.67, 19.13]	8.51	**<.001**	1.30
	British–Chinese	44.99	2.21	[39.64, 50.33]	20.34	**<.001**	3.91
	European–Chinese	30.09	2.43	[24.21, 35.97]	12.36	**<.001**	2.62
Grammar (raw scores)	British–European	1.64	0.64	[0.10, 3.17]	2.58	**.032**	0.34
	British–Chinese	10.37	0.97	[8.02, 12.72]	10.67	**<.001**	2.17
	European–Chinese	8.74	1.02	[6.26, 11.21]	8.54	**<.001**	1.83
Phonological awareness (Elision score)	British–European	−0.07	0.44	[−1.14, 0.99]	−0.17	1.00	−0.03
	British–Chinese	2.06	0.44	[0.99, 3.13]	4.63	**<.001**	0.86
	European–Chinese	2.13	0.44	[1.06, 3.20]	4.82	**<.001**	0.89
Phonological retrieval (RAN rate, digits)	British–European	0.20	0.11	[−0.05, 0.46]	1.91	.172	0.35
	British–Chinese	0.30	0.11	[0.04, 0.56]	2.82	**.016**	0.52
	European–Chinese	0.10	0.11	[−0.16, 0.35]	0.93	1.00	0.27

Note. For Vocabulary and grammar, HC3 heteroskedasticity-consistent standard errors were used due to unequal variances. *B* = unstandardized coefficient (mean difference); *SE* = standard error; CI = confidence interval; CIs and *p*-values were Bonferroni-adjusted for multiple comparisons. RAN = rapid naming. *d* = Cohen's *d*, values of 0.20, 0.50 and 0.80 indicate small, medium and large effects, respectively (Cohen, 1988). Bolded p values are statistically significant ($p < .05$).

Group Comparisons: Reading Skills

Tables 5 and 6 summarize reading and writing results. Student group was strongly associated with word-reading efficiency (SWE), text-reading rate (YAA-R), and reading comprehension (Nelson–Denny test). Pairwise comparisons showed that, on word-reading efficiency, the British group had a significant advantage over the Chinese group (large effect) but not over the European group. However, on average, the British group read texts faster than both the European and Chinese groups and obtained higher reading comprehension scores (all large effects). The European group demonstrated stronger performance than the Chinese group on all three measures (large effects for word-reading efficiency and reading comprehension; medium effect for textreading rate).

Given that the British group proved to be faster readers, and that the comprehension task was time-constrained, we examined whether the British group's advantage stemmed from the opportunity to attempt more questions within the allocated 15 min. Indeed, this was the case; there was a strong association of student group with the number of questions attempted, with the British group attempting the most questions, and the European group attempting more questions than the Chinese group. Recomputing comprehension scores as the proportion correct pet attempted questions (rather than total correct overall) removed the European group's disadvantage relative to the British group, but the Chinese group's disadvantage relative to both remained (Table 6).

Finally, we investigated whether the observed group differences in English reading efficiency and comprehension were related to individual differences in component skills. Hence, group differences were reexamined after accounting for background measures of socio-economic status, nonverbal intelligence and working memory; grammar and vocabulary as core English proficiency components; and other skills-specific components where warranted. After accounting for these covariates, the association between student group and reading measures was either substantially reduced or eliminated.

The full hierarchical regression model for word-reading efficiency explained 56.7% of variance in scores (Table 7). Stepwise analysis indicated that, after accounting for background variables (negligible effect), language proficiency (i.e., vocabulary and grammar) significantly improved model fit and explained an additional 38.9% of variance, with phonological skills (i.e., elision and RAN rate) contributing a further 13.3%. The effect of student group remained significant but was reduced to 4.1% of explained variance (down from 31.1% when it was the sole predictor; Table 5). Importantly, after

Table 5 Reading and writing performance by group

Measure	Group	Mean	*SD*	Min	Max	*F*(2,174)	*p*	R^2	f^2
Word-reading efficiency (SWE score)	British	90.49	9.41	60	104	39.26	**<.001**	.311	0.45
	European	90.82	8.21	66	104				
	Chinese	77.95	9.22	56	98				
Text-reading rate (words per minute)	British	222.66	77.47	99.73	468.57	27.08	**<.001**	.237	0.31
	European	174.38	66.19	53.97	388.42				
	Chinese	132.96	51.49	46.20	283.85				
Reading comprehension (raw score)	British	25.42	5.99	13	35	66.57	**<.001**	.433	0.76
	European	20.37	6.85	10	37				
	Chinese	12.22	5.82	3	28				
Reading comprehension: attempted questions	British	30.80	6.03	18	38	67.42	**<.001**	.437	0.78
	European	25.93	7.32	13	38				
	Chinese	16.76	6.48	3	33				
Reading comprehension: proportion correct	British	0.82	0.10	0.46	0.97	10.21	**<.001**	.109	0.12
	European	0.78	0.12	0.55	1.00				
	Chinese	0.72	0.14	0.36	1.00				

(*Continued*)

Table 5 (Continued)

Measure	Group	Mean	*SD*	Min	Max	*F*(2,174)	*p*	R^2	f^2
Orthographic knowledge	British	0.94	0.79	0	2.86	13.74	**<.001**	.153	0.18
(Spelling error rate)	European	1.36	1.31	0	7.10				
	Chinese	2.40	2.00	0	9.09				
Text-writing efficiency	British	161.05	33.19	97	244	26.71	**<.001**	.235	0.31
(text length, N of words)	European	148.85	39.68	83	236				
	Chinese	114.67	33.28	51	185				
Summarization quality	British	8.37	2.65	3	15	41.44	**<.001**	.323	0.48
(N of summarized points)	European	6.97	2.54	2	13				
	Chinese	4.31	2.13	0	12				

Note. nBritish = 59, nEuropean = 60, nChinese = 58. Attempted questions = the number of attempted reading comprehension questions in 15 min; proportion correct = the proportion of correct answers out of the total attempted reading comprehension questions. For proportion correct and spelling error rate, HC3 heteroskedasticity-consistent standard errors were used for inference due to unequal variances, and Wald F statistics is reported. f^2 = Cohen's f^2, values of 0.02, 0.15 and 0.35 indicate small, medium and large effects, respectively (Cohen, 1988). Bolded p values are statistically significant ($p < .05$).

Table 6 Group contrasts on reading and writing skills

Measure	Comparison	*B*	*SE*	95% CI	*t*	p_{bonf}	*d*
Word-reading efficiency (SWE scores)	British–European	−0.33	1.64	[−4.29, 3.64]	−0.20	1.00	−0.04
	British–Chinese	12.54	1.66	[8.54, 16.55]	7.57	**<.001**	1.40
	European–Chinese	12.87	1.65	[8.88, 16.86]	7.80	**<.001**	1.40
Text-reading rate (words per minute)	British–European	48.27	12.10	[19.03, 77.52]	3.99	**<.001**	0.73
	British–Chinese	89.69	12.20	[60.20, 119.19]	7.35	**<.001**	1.36
	European–Chinese	41.42	12.15	[12.05, 70.79]	3.41	**.002**	0.63
Reading comprehension (raw score)	British–European	5.06	1.14	[2.29, 7.82]	4.42	**<.001**	0.81
	British–Chinese	13.20	1.15	[10.41, 15.99]	11.44	**<.001**	2.12
	European–Chinese	8.14	1.15	[5.37, 10.92]	7.09	**<.001**	1.30
Reading comprehension: attempted questions	British–European	4.86	1.22	[1.92, 7.80]	4.00	**<.001**	0.73
	British–Chinese	14.04	1.23	[11.07, 17.00]	11.44	**<.001**	2.12
	European–Chinese	9.17	1.22	[6.22, 12.13]	7.51	**<.001**	1.38
Reading comprehension: proportion correct	British–European	0.04	0.02	[−0.01, 0.09]	2.03	.130	0.34
	British–Chinese	0.10	0.02	[0.05, 0.16]	4.51	**<.001**	0.85
	European–Chinese	0.06	0.02	[0.00, 0.12]	2.53	**.037**	0.50

(*Continued*)

Table 6 (Continued)

Measure	Comparison	*B*	*SE*	95% CI	*t*	p_{bonf}	*d*
Orthographic knowledge (Spelling error rate)	British–European	−0.42	0.20	[−0.90, 0.65]	−2.09	.114	−0.29
	British–Chinese	−1.46	0.28	[−2.15, −0.77]	−5.12	**<.001**	−1.01
	European–Chinese	−1.04	0.32	[−1.80, −0.28]	−3.30	**.003**	−0.72
Text-writing efficiency (text length, N of words)	British–European	12.20	6.52	[−3.56, 27.96]	1.87	.189	0.34
	British–Chinese	46.38	6.57	[30.49, 62.27]	7.06	**<.001**	1.30
	European–Chinese	34.18	6.55	[18.35, 50.00]	5.22	**<.001**	0.96
Summarization quality (N of summarized points)	British–European	1.41	0.45	[0.32, 2.49]	3.13	**.006**	0.57
	British–Chinese	4.06	0.45	[2.97, 5.16]	8.97	**<.001**	1.66
	European–Chinese	2.66	0.45	[1.57, 3.75]	5.89	**<.001**	1.08

Note. For reading comprehension proportion correct and spelling error rate, HC3 heteroskedasticity-consistent standard errors were used for inference due to unequal variances. B = unstandardized coefficient (mean difference); SE = standard error; CI = confidence interval; CIs and p-values were Bonferroni-adjusted for multiple comparisons. d = Cohen's d, values of 0.20, 0.50 and 0.80 indicate small, medium and large effects, respectively (Cohen, 1988). Bolded p values are statistically significant ($p < .05$).

Table 7 Model summaries of hierarchical regressions

Outcome variable	Model	R^2	Adj. R^2	ΔR^2	ΔR^2 p	ΔR^2 f^2
Word-reading efficiency	M1 Background	.004	−.014	.004	.889	0.00
(SWE scores)	M2 + Language	.393	.375	.389	**<.001**	0.64
	M3 + Phonological skills	.526	.505	.133	**<.001**	0.28
	M4 + Group	.567	.543	.041	**<.001**	0.09
Text-reading rate	M1 Background	.002	−.015	.002	.924	0.00
(words per minute)	M2 + Language	.320	.299	.317	**<.001**	0.47
	M3 + Word-reading efficiency	.334	.310	.015	.086	0.02
	M4 + Group	.353	.321	.018	.185	0.03
Reading comprehension	M1 Background	0.015	−0.003	.029	.478	0.03
(raw scores)	M2 + Language	0.649	0.638	0.634	**<.001**	1.81
	M3 + Text-reading rate	0.708	0.698	0.059	**<.001**	0.20
	M4 + Group	0.709	0.695	0.001	.713	0.00

(*Continued*)

Table 7 (Continued)

Outcome variable	Model	R^2	Adj. R^2	ΔR^2	ΔR^2 p	ΔR^2 f^2
Orthographic knowledge	M1 Background	0.012	−0.005	0.012	.324	0.01
(Spelling error rate)	M2 + Language	0.305	0.284	0.293	**<.001**	0.42
	M3 + Group	0.305	0.276	0.000	.959	0.00
Text-writing efficiency (Text	M1 Background	0.028	0.011	0.028	.181	0.03
length, N of words)	M2 + Language	0.302	0.281	0.274	**<.001**	0.39
	M3 + Group	0.318	0.290	0.017	.135	0.02
Summarization quality	M1 Background	0.009	−0.009	0.009	.673	0.01
(N of summarized points)	M2 + Language	0.339	0.319	0.330	**<.001**	0.50
	M3 + Writing components	0.525	0.505	0.186	**<.001**	0.39
	M4 + Group	0.540	0.514	0.015	.078	0.03

Note. Background = matrix reasoning, digit backward span, and parental education; language = vocabulary and grammar; phonological skills = elision and rapid naming rate; writing components: text length and spelling error rate; group = British, European and Chinese. R^2 = proportion of variance explained by the model; ΔR^2 = change in explained variance from the previous step. f^2 = Cohen's f^2, values of 0.02, 0.15 and 0.35 indicate small, medium and large effects, respectively (Cohen, 1988). Bolded p values are statistically significant ($p < .05$).

other predictors were included, differences between the British and European groups versus the Chinese group were no longer significant (see Appendix S1 in the Supporting Information online, Table S1.3), and RAN rate and vocabulary showed the strongest association with word-reading efficiency (see Table S1.2 in the Supporting Information online). This suggests that a smaller vocabulary and slower phonological retrieval rates underlie the Chinese group's weaker performance on word-reading efficiency. Finally, when matched on the component skills, the European group was predicted to significantly outperform the British group on this measure.

For text-reading rate, the final model explained 35.3% of variance (Table 7). As with word-reading efficiency, adding language proficiency (i.e., vocabulary and grammar) after background variables significantly improved model fit, explaining an additional 31.7% of variance. Word-reading efficiency contributed a further 1.5%, though this narrowly missed being significant. Once covariates were included, student group accounted for only 1.8% of variance (down from 23.7% when entered alone; Table 5), and no pairwise differences remained (see Appendix S2 in the Supporting Information online, Table S2.3). Among individual predictors, vocabulary was the strongest correlate of reading rate and the only significant unique predictor (see Table S2.2 in the Supporting Information online), suggesting that variation in text-reading rate across groups was primarily driven by vocabulary knowledge.

The model for reading comprehension explained 70.9% of variance. Background variables again contributed negligibly, while language proficiency added 63.4% and text-reading rate a further 5.9%. When added last, student group accounted for only 0.1% of variance (down from 43.3% as a sole predictor), and the effect was no longer significant, nor were any pairwise differences significant (see Appendix S3 in the Supporting Information online, Table 3.3). In the final model, vocabulary and reading speed remained the strongest independent predictors (see Table S3.2 in the Supporting Information online), indicating that both uniquely supported comprehension and accounted for the observed group differences.

Group Comparisons: Writing Skills

Text length, summarization quality (i.e., number of accurately summarized points), and spelling error rate served as indices of writing skill. Student group was significantly associated with all three measures, with large overall effects (Table 5). Pairwise comparisons (Table 6) showed that, relative to the British and European groups, the Chinese group produced shorter texts of lower summarization quality (large effects). They also exhibited higher spelling error

rates (medium to large effects). Differences between the European and British groups were smaller, with only summarization quality reaching significance (medium effect).

Group differences were then reexamined using hierarchical regression models to assess incremental contributions of relevant component skills (Table 7). For spelling error rate, the final model explained 30.5% of variance. Background variables accounted for only 1.2% of variance, language proficiency added 29.3%, and student group contributed none when entered last (vs. 15.3% as a sole predictor; Table 5). In the final model, grammar showed the strongest independent association with spelling error rate (see Appendix S4 in the Supporting Information online, Table S4.2).

The model for text length explained 31.8% of variance. Background variables again contributed negligibly, while language proficiency added 27.4%. After controlling for component skills, student group explained only 1.7% of variance (down from 23.5% as a sole predictor), and no pairwise differences remained significant (see Appendix S5 in the Supporting Information online, Table S5.3). In the final model, vocabulary was the only significant independent predictor (see Table S5.2 in the Supporting Information online), indicating a similar role for lexical knowledge in writing efficiency as in reading efficiency.

For summarization quality, the final model explained 54.0% of variance. Nonlinguistic background variables made no contribution, whereas core language proficiency (i.e., vocabulary and grammar) significantly improved the model fit, explaining an additional 33.0% of variance, and writing-specific components (i.e., text length and spelling error rate) adding 18.6%. When entered last, student group explained only 1.5% (down from 15.3% as a sole predictor). Although the Chinese group appeared to perform below the British group (see Appendix S6 in the Supporting Information online, Table S6.2), this difference was not significant after correction for multiple comparisons (see Table S6.3 in the Supporting Information online). In the final model, text length was the only significant independent predictor.

In sum, participants with stronger lexical and grammatical skills, regardless of group, produced longer texts and with fewer spelling errors within the task's time limits. Those who wrote more also conveyed more relevant content, underscoring the importance of vocabulary and grammar for engaging in degree-level literacy tasks. The results further highlight the impact of writing efficiency (i.e., the ability to produce text at pace) on group disparities. The association between text length and summarization quality suggests that groups differed not in the accuracy of their summaries, but rather in how many ideas

they could express within 10 min, implying that group differences in summarization quality might have been mitigated by removing time constraints.

Discussion

This study compared the English language and literacy skills of three groups of first-year undergraduates in the context of higher education in the United Kingdom: domestic students whose L1 was English, international students whose L1 was one of the European languages, and international students whose L1 was a variety of Chinese. These groups were selected because of known disparities in academic outcomes that exist between them: British and European students perform comparably, but both, on average, achieve higher academic outcomes than Chinese students. The study investigated whether these patterns reflect differences in literacy skills essential for academic success (RQ 1), and whether any group differences in literacy are explained by the component skills, particularly core English proficiency (i.e., vocabulary and grammar; RQ 2).

In relation to RQ 1, the British group showed a clear advantage over the Chinese group across all linguistic measures. The largest group difference was in vocabulary size, but robust disparities were also evidenced in grammar, phonology, and all literacy skills. On average, the Chinese group read 40% more slowly and scored 50% lower in reading comprehension. In written summaries, they produced 30% shorter texts that contained more spelling errors and 50% fewer content points than the British group. The magnitude of the differences observed between first-year Chinese and British undergraduates aligns closely with those reported by Trenkic and Warmington (2019) for Chinese postgraduates.

Crucially, our study demonstrates that such large gaps do not generalize to all international students who speak English as a L2. Compared with the European group, the British group showed a much smaller advantage in their vocabulary size, grammar knowledge, text-reading rate, reading comprehension, and written summarization. Statistically, the two groups performed indistinguishably on measures of phonology, word-reading efficiency, writing efficiency, and spelling. In fact, the gap was far greater between the Chinese and European groups than between the European and British groups. Although earlier work hinted at such differences in perception (Senyshyn et al., 2000; Trice, 2003) and vocabulary (Larson, 2017), this is the first study to demonstrate them across a comprehensive set of linguistic measures linked to academic performance while also controlling for nonverbal cognition and

socio-economic variables. These findings highlight heterogeneity among international students and caution against generalizing from one subgroup to all.

Finally, and most importantly, in regards to RQ 2, the study shows that, once individual differences in the linguistic components underpinning reading and writing were controlled for, group effects on reading comprehension, text-reading rate, and written summarization disappeared. Vocabulary, a well-established indicator of overall language proficiency, was the strongest correlate of these literacy skills and the main driver of group disparities.

The link between core language proficiency and literacy skills in this population is theoretically and practically important for two reasons. First, text-level literacy depends on more than core language proficiency, also drawing on general cognitive abilities and broader sociocultural knowledge. Our results show that literacy differences between Chinese, European and British students cannot be attributed to cognitive ability. They also indirectly rule out sociocultural influences by fully accounting for literacy performance through English proficiency. Second, given selective university recruitment based on minimum language requirements, some could expect variation in English proficiency between local and international students to be minimal and thus inconsequential for literacy. Yet, our findings indicate that many students begin university with proficiency levels that constrain the literacy skills essential for academic success. This challenges the assumption that meeting minimum entry requirements ensures linguistic readiness and that students need support with only academic English and disciplinary literacy norms (Jenkins & Wingate, 2015). Although such support remains important, our results suggest that the ability to develop and efficiently deploy complex literacy skills in English is unevenly distributed, with vocabulary size being a critical limiting element for some.

The Interplay Between Reading Speed, Reading Comprehension, and Vocabulary

Research on the linguistic predictors of academic outcomes in higher education has rarely focused on text-reading efficiency. However, our results indicate that this is a dimension of literacy whereupon students studying in their L1 may have particular advantage (see also Babayiğit & Trenkic, 2025; Siegelman et al., 2024). For instance, the British group in our study read silently at an average rate of 223 words per minute, which is close to the estimated silent reading rate (238 words per minute) of adult native speakers of English (Brysbaert, 2019). In comparison, the European and Chinese groups read at 174 and 133 words per minute, respectively. This is important to note because faster

reading improves one's capacity to read more extensively within the available study time, thus enabling more learning in preparation for assessment.

Furthermore, faster reading can confer meaningful advantages in assessment itself, particularly in timed exams. The results from our reading comprehension test speak to this point. As part of this task, under exam-like time constraints, participants read seven texts, each followed by multiple-choice comprehension questions. The British group scored highest but, importantly, this advantage stemmed from the fact that, as faster readers, they attempted more comprehension questions than the other two groups. When scores were adjusted for the proportion of correct answers relative to attempted questions, the British group advantage over the European group disappeared and was substantially reduced relative to the Chinese group. This finding has two major implications: First, it demonstrates that reading speed could be an important contributor to the persistent degree-awarding gaps between domestic and international students; second, it supports arguments for untimed assessments as a more equitable format (Gernsbacher et al., 2020).

Importantly, English vocabulary size was strongly associated with both reading speed and reading comprehension. In fact, controlling for individual differences in vocabulary eliminated group disadvantages in both reading speed and reading comprehension. Beyond confirming its central role in reading (Jeon & Yamashita, 2014; Perfetti & Stafura, 2014; Siegelman et al., 2024), this finding provides crucial evidence that, for some higher education students in United Kingdom, the number of known English words substantially limits both the speed at which they can read and how much they can understand.

Furthermore, the interplay between vocabulary, reading speed, and reading comprehension documented in our study provides new insights for debates on how to define adequate vocabulary size in tertiary education. A target of eight to nine thousand word families is often recommended for students entering English-medium higher education (Miralpeix, 2025), based on research suggesting that this level provides approximately 98% coverage of running words in a wide range of written texts (Nation, 2006).[2] Because the students in the Chinese group had an average knowledge of 7.4 thousand word families, our study helps confirm that vocabulary knowledge below the eight to nine thousand threshold constrains both reading speed and comprehension. By contrast, the European group's results, having an average knowledge of 11.4 thousand word families, suggest that this level of lexical knowledge—though still well below that of the British group (16.4 thousand)—can support L1-comparable comprehension, albeit less efficiently.

The European group's performance aligns with evidence that advanced L2 English speakers can attain native-like levels of reading comprehension despite residual gaps in component skills (Siegelman et al., 2024), and that parity in comprehension is more common than parity in reading rate (Kuperman et al., 2025). In our data, however, text-reading rate (and not just reading comprehension) was strongly associated with vocabulary size. Given that the European group matched, or closely approximated, the British group in grammatical, phonological, and word-decoding skills essential for fluent reading, their slower pace likely reflects encountering more unfamiliar words and inferring meaning from context, making comprehension more time-demanding and effortful (cf. Pellicer-Sánchez et al., 2024). For the Chinese group, with lower overall proficiency, additional time alone could not fully offset the impact of a substantially smaller vocabulary on comprehension.

Academic Writing

In the written summarization task, the British group produced the longest texts and covered the most content points from the source passage. Summarizing is a key academic skill, central to coursework assessment and strongly linked to grades (Trenkic & Warmington, 2019), and, like reading, summarizing is strongly underpinned by language proficiency. Once vocabulary and grammar were controlled for, group effects on both text length and content coverage disappeared. Vocabulary was the strongest predictor of text length, which, in turn, predicted the number of content points summarized. This parallels the reading findings, underscoring the central role of vocabulary in academic performance and showing how time constraints can amplify disparities rooted in linguistic proficiency.

Implications, Limitations and Future Direction

This study shows that, within higher education in the United Kingdom, core English proficiency—particularly vocabulary size—critically shapes students' reading speed, reading comprehension, and written summarization, all of which are fundamental to academic success. The substantial group differences suggest that current language entry standards (Pearson, 2020) may be contributing to the persistent degree-awarding gaps between domestic and international students. Although these minimum requirements reflect broader institutional and financial considerations, our findings highlight the need to inform applicants not only of the minimum English proficiency accepted for admission, but also of the levels needed for them to perform at their full academic potential.[3]

The findings also help explain why degree-awarding gaps are not observed for European students (Borrett & Foster, 2023). Our results suggest that, although they enter university with slightly lower English proficiency than their British peers, European students typically possess the language mastery necessary to sustain comparable levels of reading comprehension and written summarization and thus achieve similar academic performance. This supports the view that native-like English proficiency is not a prerequisite for academic success; however, a threshold level is required to ensure that language does not impede learning or assessment. The contrast between the European and Chinese groups further indicates that, when language entry requirements are set below recommended proficiency standards, the linguistic distance of students' L1 from English can affect the linguistic readiness that they bring to their studies.

Although the findings align with evidence that linguistic distance affects the time and effort required to reach advanced L2 proficiency, closeness does not guarantee high proficiency, nor does distance preclude it. Support should target proficiency differences regardless of L1 background. We also acknowledge that, although linguistic distance provides the most theoretically parsimonious explanation for the observed differences in English proficiency, groups may also differ in their exposure to English, teaching practices, or culturally specific approaches to test preparation (Trenkic & Hu, 2021). Future research should disentangle these variables in order to inform targeted interventions that aim to enhance L2 students' linguistic readiness.

The inclusion of widely used measures of vocabulary size and reading rate offers reassurance that the samples were broadly representative. However, in order to establish the true extent to which particular student groups begin their studies with language skills that place them at a systematic disadvantage, sector-level data is needed. At the policy level, our findings highlight that data regarding international students' language background and English proficiency is the type of evidence needed to facilitate a more meaningful interpretation of group disparities in academic outcomes and support evidence-based efforts to reduce them.

Conclusion

This study underscores the critical role of vocabulary size and processing efficiency in supporting the complex literacy skills essential for academic success in tertiary education. Specifically, we show that, for some higher education students in the United Kingdom, even though selected for language proficiency, limited English vocabulary significantly constrains skills integral to effective learning and performance (e.g., reading speed, reading comprehension, and the

ability to summarize texts in writing). These findings suggest that vocabulary-driven differences may be an important contributor to the large and persistent degree-awarding gaps between international and domestic tertiary education students in the United Kingdom.

Crucially, our findings highlight that international students are a diverse population and that the collective label often conceals substantial disparities in English proficiency across subgroups. By comparing European, Chinese, and British students, this study shows that the large language and literacy gaps reported in previous research are neither inevitable nor generalizable to all international students. For example, the European students in our sample entered university with English language and literacy skills much closer to those of their British peers than to their Chinese counterparts. This suggests that, when admission language requirements fall below the recommended thresholds for academic study, students from typologically distant language backgrounds are more likely to start with lower proficiencies and, consequently, face greater challenges. These findings underscore the need for equitable preparation and ongoing language support to enable all students to realise their full academic potential.

Final revised version accepted 19 February 2026

Open Research Badges

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data are available at https://osf.io/rdn6s/.

Notes

1 The IELTS scale in the relevant range corresponds to: 4 = *limited user*, 5 = *modest user*, 6 = *competent user*, 7 = *good user*, 8 = *very good users*, 9 = *expert user*. For further descriptors, see IELTS (2024).
2 However, some analyses of actual assigned texts suggest that some require up to 20 thousand word families for full coverage (Vicary & Treffers-Daller, 2024).
3 Expressed as an IELTS score, the recommended level for linguistically demanding academic programmes is 7.5 or above, with half a band lower (7.0) acceptable for less linguistically demanding programmes (IELTS, 2024).

References

August, D., & Shanahan, T. (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth*. Routledge.

Babayiğit, S., & Trenkic, D. (2025). Vocabulary knowledge is key to understanding and addressing disparities in higher education. *British Journal of Educational Psychology*. https://doi.org/10.1111/bjep.70000

Borrett, A., & Foster, P. (2023). Non-EU international students at UK universities less likely to get top grades, data shows. *Financial Times*. https://www.ft.com/content/30217f1d-257f-4826-8aba-ec9329a4ab4d

Brown, J. I., Fishco, V. V., & Hanna, G. S. (1993). *Nelson-Denny reading test (Forms G & H)*. Riverside.

Brysbaert, M. (2019). How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language*, *109*, 104047. https://doi.org/10.1016/j.jml.2019.104047

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Council of Europe. (2025). Global scale – Table 1 (CEFR 3.3): Common reference levels. https://www.coe.int/en/web/common-european-framework-reference-languages/table-1-cefr-3.3-common-reference-levels-global-scale

Crawford, I., & Wang, Z. (2015). The impact of individual factors on the academic attainment of Chinese and UK students in higher education. *Studies in Higher Education*, *40*(5), 902–920. https://doi.org/10.1080/03075079.2013.851182

Daller, M. H., Kuiken, F., Trenkic, D., & Vedder, I. (2021). Linguistic predictors of academic achievement amongst international students and home students in higher education: Introduction. *International Journal of Bilingual Education and Bilingualism*, *24*(10), 1453–1457. https://doi.org/10.1080/13670050.2021.1909452

Daller, M. H., & Phelan, E. (2013). Predicting international student study success. *Applied Linguistics Review*, *4*(1), 173–193. https://doi.org/10.1515/applirev-2013-0008

Eddey, P., & Baumann, C. (2011). Language proficiency and academic achievement in postgraduate business degrees. *International Education Journal: Comparative Perspectives*, *10*(1), 34–46.

Gernsbacher, M. A., Soicher, R. N., & Becker-Blease, K. A. (2020). Four empirically based reasons not to administer time-limited tests. *Translational Issues in Psychological Science*, *6*(2), 175–190. https://doi.org/10.1037/tps0000232.

Harrington, M., & Roche, T. (2014). Identifying academically at-risk students in an English-as-a-Lingua-Franca university setting. *Journal of English for Academic Purposes*, *15*, 37–47. https://doi.org/10.1016/j.jeap.2014.05.003

Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, *177*, 263–277. https://doi.org/10.1016/j.cognition.2018.04.007.

Hayes, J. R. (2012). Modeling and remodeling writing. *Written Communication*, *29*, 369–388. https://doi.org/10.1177/0741088312451260

HESA. (2025a). *HE qualifiers by location of HE provider, mode of study, level of qualification obtained, permanent address, sex, classification of first degree and academic year*. Higher Education Statistics Agency. https://www.hesa.ac.uk/data-and-analysis/students/table-18

HESA. (2025b). *HE student enrolments by permanent address 2019/20 to 2023/24*. Higher Education Statistics Agency. https://www.hesa.ac.uk/data-and-analysis/sb271/figure-9

Hu, R., & Trenkic, D. (2021). The effects of coaching and repeated test-taking on Chinese candidates' IELTS scores, their English proficiency, and subsequent academic achievement. *International Journal of Bilingual Education and Bilingualism*, *24*, 1486–1501. https://doi.org/10.1080/13670050.2019.1691498

Hyland, K. (2016). Academic publishing and the myth of linguistic injustice. *Journal of Second Language Writing*, *31*, 58–69. https://doi.org/10.1016/j.jslw.2016.01.005

Iannelli, C., & Huang, J. (2014). Trends in participation and attainment of Chinese students in UK higher education. *Studies in Higher Education*, *39*(5), 805–822. https://doi.org/10.1080/03075079.2012.754863

IELTS. (2024). *Guide for organisations*. Cambridge University Press and Assessment. https://s3.eu-west-2.amazonaws.com/ielts-web-static/production/Guides/ielts-guide-for-institutions.pdf

Jenkins, J., & Wingate, U. (2015). Staff and students' perceptions of English language policies and practices in international universities: a case study from the UK. *Higher Education Review*, *47*, 47–73.

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning*, *64*(1), 160–212. https://doi.org/10.1111/lang.12034

Kojima, M., In'nami, Y., & Kaneta, T. (2022). L2 writing and its external correlates: A meta-analysis. In E. H. Jeon & Y. In'nami (Eds.) *Understanding L2 proficiency: Theoretical and meta-analytic investigations* (pp. 159–211). John Benjamins Publishing Company.

Kuperman, V. (2025). How does language distance affect reading fluency and comprehension in English as second language? *Studies in Second Language Acquisition*, *47*, 757–773. https://doi.org/10.1017/S027226312510065X

Kuperman, V., Schroeder, S., Acartürk, C., Agrawal, N., Alexandre, D. M., Bolliger, L. S., Brasser, J., Campos-Rojas, C., Drieghe, D., Durdevic, D. F., Gadelha de Freitas, L. V. G., Goldina, S., Ibanez Orellana, R., Jager, L. A., Johannesson, O. I., Khare, A., Kharlamov, N., Knudsen, H. B. S., Kristjansson, A., … Siegelman, N. (2025). New data on text reading in English as a second language: The Wave 2 expansion of the Multilingual Eye-Movement Corpus (MECO). *Studies in Second Language Acquisition*, *47*(2), 677–695. https://doi.org/10.1017/S0272263125000105

Larson, M. (2017). *Thresholds, text coverage, vocabulary size, and reading comprehension in applied linguistics* [Doctoral dissertation, Te Herenga Waka-Victoria University of Wellington]. Open Access Te Herenga Waka-Victoria University of Wellington. https://doi.org/10.26686/wgtn.17057906.v1

Lenth, R. V. (2025). *emmeans: Estimated marginal means aka least-squares means* (R Package; Version 1.11.2-8). Retrieved from https://cran.rproject.org/web/packages/emmeans/emmeans.pdf

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

McCarron, S. P., & Kuperman, V. (2022). Effects of year of post-secondary study on reading skills for L1 and L2 speakers of English. *Journal of Research in Reading*, *45*, 43–64. https://doi.org/10.1111/1467-9817.12380

Miralpeix, I. (2025). Vocabulary size in a second language. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (2nd ed). Wiley Blackwell. https://doi.org/10.1002/9781405198431.wbeal1270.pub2

Morrison, J., Merrick, B., Higgs, S., & Le Métais, J. (2005). Researching the performance of international students in the UK. *Studies in Higher Education*, *30*(3), 327–337. https://doi.org/10.1080/03075070500095762

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, *63*(1), 59–82. https://doi.org/10.3138/cmlr.63.1.59

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*, 9–13. https://doi.org/10.26686/wgtn.12552197

Nisbet, K., Bertram, R., Erlinghagen, C., Pieczykolan, A., & Kuperman, V. (2022). Quantifying the difference in reading fluency between L1 and L2 readers of English. *Studies in Second Language Acquisition*, *44*, 407–434. https://doi.org/10.1017/S0272263121000279

Oliver, R., Vanderford, S., & Grote, E. (2012). Evidence of English language proficiency and academic achievement of non-English-speaking background students. *Higher Education Research & Development*, *31*(4), 541–555. https://doi.org/10.1080/07294360.2011.653958.

Pearson, W. S. (2020). Mapping English language proficiency cut-off scores and pre-sessional EAP programmes in UK higher education. *Journal of English for Academic Purposes 45*, 100866. https://doi.org/10.1016/j.jeap.2020.100866

Pellicer-Sánchez, A., Webb, S., & Wang, A. (2024). How does lexical coverage affect the processing of L2 texts? *Applied Linguistics*, *45*(6), 953–972. https://doi.org/10.1093/applin/amae062

Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, *18*, 22–37. https://doi.org/10.1080/10888438.2013.827687

R Core Team. (2025). *R: A language and environment for statistical computing* (v4.5.) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Roche, T., & Harrington, M. (2013). Recognition vocabulary knowledge as a predictor of academic performance in an English as a foreign language setting. *Language Testing in Asia*, *3*, 1–13. https://doi.org/10.1186/2229-0443-3-12

Schepens, J., Van Hout, R., & Jaeger, T. F. (2020). Big data suggest strong constraints of linguistic similarity on adult language learning. *Cognition*, *194*, 104056. https://doi.org/10.1016/j.cognition.2019.104056

Schmidtke, D., Lu, G., & Moro, A. L. (2025). EAL reading growth during and beyond an undergraduate transition program: A 4-year longitudinal eye-movement study. *Reading and Writing*. https://doi.org/10.1007/s11145-025-10664-6

Schmidtke, D., & Moro, A. L. (2021). Determinants of word-reading development in English learner university students: A longitudinal eye movement study. *Reading Research Quarterly*, *56*(4), 819–854. https://doi.org/10.1002/rrq.362

Schmidtke, D., Rahmanian, S., & Moro, A. L. (2023). Tracking reading development in an English language university-level bridging program: Evidence from eye-movements during passage reading. *Bilingualism: Language and Cognition*, *26*, 356–370. https://doi.org/10.1017/S1366728922000542

Schoonen, R., Snellings, P., Stevenson, M., & Van Gelderen, A. (2009). Towards a blueprint of the foreign language writer: The linguistic and cognitive demands of foreign language writing. In R. Manchón (Ed.), *Writing in foreign language contexts: Learning, teaching, and research* (pp. 77–101). Multilingual Matters. https://repository.ubn.ru.nl/handle/2066/191571

Senyshyn, R. M., Warford, M. K., & Zhan, J. (2000). Issues of adjustment to higher education: International students' perspectives. *International Education*, *30*(1), 1–17. https://www.proquest.com/docview/1311677222?fromopenview=true&imgSeq=1&pq-origsite=gscholar&sourcetype=Scholarly%20Journals

Siegelman, N., Elgort, I., Brysbaert, M., Agrawal, N., Amenta, S., Arsenijević Mijalković, J., Chang, C. S., Chernova, D., Chetail, F., Clarke, A. J. B., Content, A., Crepaldi, D., Davaabold, N., Delgersuren, S., Deutsch, A., Dibrova, V., Drieghe, D., Durdevic, D. F., Finch, B., … Kuperman, V. (2024). Rethinking first language–second language similarities and differences in English proficiency: Insights from the ENglish Reading Online (ENRO) project. *Language Learning*, *74*(1), 249–294. https://doi.org/10.1111/lang.12586

Spit, S., Andringa, S., & Ryan, O. (2025). Investigating the relation between second language proficiency and study success using a causal inference approach. *Language Learning*. https://doi.org/10.1111/lang.12713

Thorpe, A., Snell, M., Davey-Evans, S., & Talman, R. (2017). Improving the academic performance of non-native English-speaking students: The contribution of pre-sessional English language programmes. *Higher Education Quarterly*, *71*(1), 5–32. https://doi.org/10.1111/hequ.12109

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *Test of word reading efficiency* (2nd ed.) (TOWRE-2). Pro-Ed.

Trenkic, D., & Hu, R. (2021). Teaching to the test: The effects of coaching on English-proficiency scores for university entry. *Journal of the European Second Language Association*, *5*(1), 1–15. https://doi.org/10.22599/jesla.74

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

Trice, A. G. (2003). Faculty perceptions of graduate international students: The benefits and challenges. *Journal of Studies in International Education*, *7*(4), 379–403. https://doi.org/10.1177/1028315303257120

Vicary, A., & Treffers-Daller, J. (2024). Reading against all odds: Creative coping strategies of international students in HE. *International Journal of Bilingual Education and Bilingualism*, *27*(8), 1128–1141. https://doi.org/10.1080/13670050.2024.2345698

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive test of phonological processing*: CTOPP. Pro-Ed.

Warmington, M., Stothard, S. E., & Snowling, M. J. (2013). Assessing dyslexia in higher education: The York Adult Assessment Battery – Revised. *Journal of Research in Special Educational Needs*, *13*(1), 48–56. https://doi.org/10.1111/j.1471-3802.2012.01264.x.

Wechsler, D. (2010). *Wechsler adult intelligence scale - Fourth UK edition* (WAIS-IV).

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Accessible Summary

Appendix S1. Word-Reading Efficiency (SWE Score).

Appendix S2. Text-Reading Rate (Words Per Minute).

Appendix S3. Reading Comprehension (Raw Score).

Appendix S4. Orthographic Knowledge (Spelling Error Rate).

Appendix S5. Text-Writing Efficiency (Text Length).

Appendix S6. Summarization Quality (Number of Accurately Summarized Content Points).

Appendix S7. Sensitivity Power Analyses.