ETS TOEFL

Quality Beyond Measure.

TOEFL® Research Report
TOEFL–RR-91
ETS Research Report No. RR–21-04

Researching Academic Reading in Two Contrasting English as a Medium of Instruction Contexts at a University Level

Nathaniel Owen

Prithvi N. Shrestha

Anna Kristina Hultgren

December 2021

The *TOEFL*® test is the world's most widely respected English language assessment, used for admissions purposes in more than 130 countries including Australia, Canada, New Zealand, the United Kingdom, and the United States. Since its initial launch in 1964, the TOEFL test has undergone several major revisions motivated by advances in theories of language ability and changes in English teaching practices. The most recent revision, the *TOEFL iBT*® test, contains a number of innovative design features, including integrated tasks that engage multiple skills to simulate language use in academic settings and test materials that reflect the reading, listening, speaking, and writing demands of real-world academic environments. In addition to the TOEFL iBT, the TOEFL Family of Assessments has expanded to provide high-quality English proficiency assessments for a variety of academic uses and contexts. The TOEFL Young Students Series (YSS) features the *TOEFL*® *Primary*™ and *TOEFL Junior*® tests, designed to help teachers and learners of English in school settings. The *TOEFL ITP*® Assessment Series offers colleges, universities, and others an affordable test for placement and progress monitoring within English programs.

Since the 1970s, the TOEFL tests have had a rigorous, productive, and far-ranging research program. ETS has made the establishment of a strong research base a consistent feature of the development and evolution of the TOEFL tests, because only through a rigorous program of research can a testing company demonstrate its forward-looking vision and substantiate claims about what test takers know or can do based on their test scores. In addition to the 20-30 TOEFL-related research projects conducted by ETS Research & Development staff each year, the TOEFL Committee of Examiners (COE), composed of distinguished language-learning and testing experts from the academic community, funds an annual program of research supporting the TOEFL family of assessments, including projects carried out by external researchers from all over the world.

To date, hundreds of studies on the TOEFL tests have been published in refereed academic journals and books. In addition, more than 300 peer-reviewed reports about TOEFL research have been published by ETS. These publications have appeared in several different series historically: TOEFL Monographs, TOEFL Technical Reports, TOEFL iBT Research Reports, and TOEFL Junior Research Reports. It is the purpose of the current TOEFL Research Report Series to serve as the primary venue for all ETS publications on research conducted in relation to all members of the TOEFL Family of Assessments.

To obtain more information about the TOEFL programs and services, use one of the following:

E-mail: toefl@ets.org **Web site:** www.ets.org/toefl

ETS is an Equal Opportunity/Affirmative Action Employer.

As part of its educational and social mission and in fulfilling the organization's non-profit Charter and Bylaws, ETS has and continues to learn from and also to lead research that furthers educational and measurement research to advance quality and equity in education and assessment for all users of the organization's products and services.

TOEFL Research Report Series and ETS Research Report Series ISSN 2330-8516

RESEARCH REPORT

Researching Academic Reading in Two Contrasting English as a Medium of Instruction Contexts at a University Level

Nathaniel Owen, Prithvi N. Shrestha, & Anna Kristina Hultgren

School of Languages and Applied Linguistics, The Open University, Milton Keynes, England,

This project examined academic reading in two contrasting English as a medium of instruction (EMI) university settings in Nepal and Sweden and the unique challenges facing students who are studying in a language other than their primary language. The motivation for the project was to explore the role of high-stakes testing in EMI contexts and the implications for the design of the *TOEFL iBT*® test. We employed a sequential mixed-methods approach to gather substantive and authentic qualitative data from stakeholders immersed in EMI settings. A small sample of students (Nepal = 19, Sweden = nine) were asked to complete reading logs over a period of 3 weeks so we could determine the types of texts and reading load associated with diverse EMI settings. Additionally, a larger cohort of students from each setting (Nepal = 69, Sweden = 60) completed questionnaires examining academic reading demands, reading skills, and practices. Students who completed the questionnaires also completed the reading section of the TOEFL iBT test. The same students also completed a *TOEFL*® family of tests suitability questionnaire so we could consider the suitability of the TOEFL iBT test for EMI contexts. Following test completion, a series of semistructured interviews (Nepal = 21, Sweden = 23) focused more closely on students' perspectives of reading demands in their academic contexts and the suitability of the reading section of the TOEFL iBT test to make claims about readiness to study in EMI contexts. Our findings revealed that different EMI contexts have different standards of high and low academic reading proficiency and that these differences may occur due to differences in educational experiences of the respective cohorts. The findings offer important new insights into academic reading and assessment in EMI contexts. Students in EMI contexts are sensitive to violations of expectations regarding test-taking experiences (face validity). The study has implications for the design of test tasks, which should consider local, contextual varieties of English.

Keywords Context validity; mixed-methods study; English-medium instruction; student perceptions; academic reading; the *TOEFL iBT*® test

doi:10.1002/ets2.12317

The *TOEFL iBT*® test is being increasingly used in new contexts with English as a medium of instruction (EMI). EMI is defined as "the use of the English language to teach academic subjects (other than English itself) in countries or jurisdictions where the first language of the majority of the population is not English" (Macaro et al., 2018, p. 37). Some recent studies have examined the predictive validity of the test and demonstrated that the test is appropriate for university admissions in different EMI contexts (e.g., see O'Dwyer et al., 2018, in the Turkish context; Ockey et al., 2014, in the Japanese context). The increasing use of the test necessitates a new empirical research base to be able to make additional claims for score interpretation and use. Research involving domain analysis within EMI contexts, in conjunction with research into the relevance of the TOEFL iBT test to these contexts, will provide an evidential basis for ETS to make additional claims regarding test use of the TOEFL iBT test or clarify the extent to which the domain of reading for academic purposes is consistent across these contexts.

Within EMI contexts, exposure to English language use within and outside of the university context is presumed to be lower than English language use in an English-dominant context such as those found in the United States or the United Kingdom. Thus, academic, didactic, and pedagogic practices, as well as access to and uptake of literacy resources (print and electronic), cannot be assumed to be the same in these contexts. Additionally, significant differences in EMI practices have been documented across Global North contexts[1] (Hultgren et al., 2015; Wächter & Maiworm, 2014). Therefore, it is reasonable to hypothesize that EMI contexts will also be markedly different between Global North and Global South contexts (Dearden, 2014). For example, many Global South countries face a lack of qualified teachers to teach in EMI, unlike the Global North countries (Dearden, 2014). Domestic students (non-English primary language [L1]) are often

Corresponding author: N. Owen, E-mail: nathaniel.owen@open.ac.uk

not screened for academic English literacy, with universities only requiring high school English for admission. English proficiency may therefore be unknown, which may have an impact on students' success in their academic programs, as research has suggested that reading and learning from texts in a secondary language (L2) requires more time and effort than reading and learning in L1 (Mežek, 2013). When we compared Global North with Global South contexts in terms of English proficiency in different countries drawing on English First's (2019) publicly available English Proficiency Index 2019, differences emerged that required exploration. For example, the EF English Proficiency Index 2019 (based on 2.3 million test takers) showed that, for the countries of interest in this study, Sweden was at Number 2 and Nepal at Number 66 in the ranking of 100 countries. This direct comparison between the two countries, however, needs to be considered cautiously because there may be sample biases (e.g., test takers from EMI and non-EMI backgrounds) in the EF Standard English Test (www.efset.org) as in any language test. Despite this sample bias, Sweden seemed to have a higher English proficiency level than Nepal did. Whereas the primary purpose of the TOEFL iBT test is for admission to higher education, the use of the reading section of the TOEFL iBT test, or some other reading comprehension test, could be extended to gauge entry-level proficiency, to place students into levels, or to inform curriculum design for a pre-sessional course of English. Although ETS cautions against generalizations at the country level, the test can be used to compare students' English proficiency levels across institutions for which such admissions tests are not typically employed (ETS, 2020).

University EMI programs have started to require proof of English language proficiency upon entry, with the most common tests in use being the TOEFL iBT test and the IELTS test (Rogier, 2012; Saarinen & Nikula, 2013). However, a lack of research into differences between EMI contexts (Global North and Global South) and North American, United Kingdom, and Australian contexts, means there is no research to date that specifically investigates the validity of using these tests for university admission in EMI contexts. Using the TOEFL iBT test for admission in an academic program delivered in English in a country where English is not the native language requires a specific validation argument.

In light of these concerns, we conducted this study and focused on two key areas. First, the study consists of domain analysis within two EMI contexts, Stockholm University in Sweden and Tribhuvan University in Nepal, and the differences in the domain of *reading for academic purposes* between these contexts. Second, the research investigates whether language practices elicited by the reading section of the TOEFL iBT test align with the domain of academic reading identified in the two EMI contexts.

In order to contextualize the study further, it is worth pointing out the key differences between the two universities (further discussed in subsequent sections):

- The source of the student population in each university is different. Whereas many students at Stockholm are international, the students at Tribhuvan are local.
- In Nepal, the students in science, technology, engineering, and math (STEM) have no choice but to study in EMI programs because the courses are not available in Nepali or any local language, whereas in Sweden the students sought to study (by their choice) at Stockholm for internationalized higher education, meaning EMI.
- English is used as a lingua franca among Stockholm students as they are international students who have different language backgrounds whereas English is imposed on Nepalese students at Tribhuvan because there is no need for them to communicate in English as they all speak Nepali.

Review of Literature

EMI in Sweden

Sweden is among the top providers of EMI courses in higher education in Europe (Hultgren et al., 2015; Wächter & Maiworm, 2008, 2014). Over the last decade, Sweden has experienced an "unprecedented expansion of EMI" at university level (Airey et al., 2015, p. 562). This growth is in stark contrast to pre-tertiary levels where students are rarely exposed to EMI. According to the latest Europe-wide survey conducted in 2014, 822 university programs were conducted in EMI at Swedish universities compared to 123 in 2007 (Wächter & Maiworm, 2014), a number likely to have increased since the report was published. In 2014, 28% of all university degree programs were being offered in English (Salö & Josephson, 2014). EMI varies significantly according to disciplines, with STEM subjects and the social sciences typically leading the pace over the humanities (e.g., Kuteeva & Airey, 2013). In addition to (some) programs officially being taught in

English, much of the course literature as well as the teaching and learning resources are also typically in English, with only a small minority in Swedish (Airey & Linder, 2008; Bolton & Kuteeva, 2012). Although English is no doubt pervasive in Swedish higher education, it is worth noting that the linguistic practices tend to be significantly more multilingual (Söderlundh, 2012).

EMI in Nepal

Nepal is one of the developing economies located in South Asia in the Global South. Despite the fact that Nepal is a multilingual country with at least 123 local languages and the national language, Nepali, English has always enjoyed a higher symbolic status and value (Phyak, 2016; Sah & Li, 2018; Shrestha, 2009). Given the perceived economic value and social status attached to the English language, EMI is often limited to affluent families in the highly stratified society in Nepal (Phyak, 2016). This situation is probably perpetuated by the legitimation of EMI through an education act in 2010 (Sah & Li, 2018) and the lack of preparing teachers for EMI, as the government admits (Ministry of Education, 2016).

English language is the most widely used foreign language in the country and has gained currency in the last few decades in all spheres of life including education (Phyak, 2016; Sah & Li, 2018; Shrestha, 2009). As a subject in education, English language has been a compulsory subject from elementary school to undergraduate degree level in Nepal for over 50 years (Shrestha, 2009). The first higher education college in Nepal to teach through the medium of English was established in 1918. Pre-tertiary private schools established in Nepal throughout the twentieth century have always taught in English, although the quality of these private schools varies significantly depending on location and their resources (Shrestha, 2009). As a result, many Nepali university students, depending on their academic pathway, do not have much choice regarding the language of instruction, with the majority of academic pathways in STEM delivered through EMI. The key concern, as reported by previous EMI research around the world and a few studies in Nepal, strongly highlights the lack of qualified teachers who can teach in English. This reality is incongruent with the growing demand for more EMI schools in the country (Ministry of Education, 2016).

Research Into Reading for Academic and Testing Purposes in EMI Contexts

This research is purposefully limited to reading in English for academic purposes. Academic study tends to be difficult for individuals who are studying through the medium of English in a country for which English is not the native language. Successful reading is crucial for academic study. The importance of reading is evident in the central role that this skill occupies in the TOEFL iBT test. Moore et al. (2007) noted that despite the significant changes such as integration of language skills that have been undertaken in major language tests, including the TOEFL iBT test, over the last 30 years, all high-stakes English tests retain a dedicated section for reading.

"Readers" in the context of this study refers to cognitively mature individuals who are already literate in their first language, have been learning to read in English as a second or foreign language (Koda, 2005), and are now active readers on an academic pathway in which English is the principal medium of instruction. These readers draw from a number of L1 capabilities in L2 text engagement, including visual information extraction; incremental information integration; and text meaning, moderated by prior knowledge construction (Koda, 2005). Language learners draw from a number of L1 capabilities in L2 interaction, including knowledge of morphosyntax, phonology, pragmatics, metalinguistic awareness, and communicative strategies. L1 linguistic features influence L2 acquisition and the types of strategies that L2 learners employ. Information processing procedures vary systematically across languages (Koda, 2005), suggesting there is no universal framework of how L2 learners from different linguistic backgrounds interact with texts. L2 decoding efficiency is at least partially determined by L1 – L2 orthographic distance (Grabe, 2009; Koda, 2005). Metalinguistic awareness — deliberate and goal-oriented deployment of a strategic action (Phakiti, 2003) — can also be researched in EMI contexts if the language learners are aware of (and can report using) these strategies.

Reading in a test-taking environment is goal oriented. The purpose of reading is to extract very specific information for the purpose of addressing a particular task or set of tasks. This action is typically executed under a time constraint, unlike the target domain. Within the reading section of the TOEFL iBT tests, different reading tasks are associated with reading purpose as initially conceptualized in Enright et al. (2000) and further developed in Pearlman (2008) and Jamieson et al. (2008). The first reading purpose is *reading for basic comprehension*, which requires a reader to understand the main

ideas or the main points of the text or to form some understanding of the main theme of the text. The second is *inferencing*, which requires test takers to comprehend ideas not explicitly stated in a text, a skill that could be distinguished from basic comprehension (Pearlman, 2008, p. 242) during development of the TOEFL iBT test. *Reading to learn items* focus on developing "an organised mental representation of a text that supports the recollection and reconstruction of ideas in the text" (Jamieson et al., 2008, p. 72).[2]

Research Questions

Based on the literature, the present study investigates reading for academic purposes in two EMI contexts through a sequential mixed-methods research incorporating qualitative and quantitative methodologies, aiming to address the following overarching research questions:

- What are the reading practices of students in the two EMI contexts under investigation?
- Is the reading section of the TOEFL iBT suitable as an entry test for different EMI program courses in the two contexts?

Methodology

Overview of Research Methods and Tools

Research Question 1 considers the reading practices of students in the respective EMI contexts and is addressed through the following means:

- A detailed reading log by a self-selected sample of students from each context;
- Likert-type items from the questionnaire to university students in each of the two EMI contexts, examining past and present academic reading practices and strategies; and
- Follow-up interviews in each context, examining detailed reading practices of students within each domain.

Research Question 2 considers the suitability of the TOEFL iBT reading test for each EMI context and is addressed through the following means:

- Delivery of the reading section of the TOEFL iBT test (60-min version) to the students who completed the questionnaires at the same university in the two EMI contexts; taken under test conditions; and
- Follow-up questionnaires to the same students on their experience of doing the test and their opinion of the suitability of the TOEFL iBT test within their context.

The test data from the reading section of the TOEFL iBT test was compared across both EMI contexts for each of the three reading purposes (i.e., reading for basic comprehension, inferencing, reading to learn). The test contained three passages: one related to psychology, one to biology, and one to paleontology. This study used a research version of the TOEFL iBT test rather than an operational test. The content of the test predated the enhancements made to the test in August 2019, which resulted in a shorter reading test (see Papageorgiou et al., 2019, for further details of the revisions).

Study Participants

In total, 129 students participated in the study: 69 students in Nepal and 60 in Sweden. Brief descriptive data for the cohort is outlined in Table 1.

The two cohorts were similar in age, although the Swedish cohort exhibited slightly more variation than the Nepalese cohort due to participation by several mature students. Males were overrepresented in Nepal, whereas females were overrepresented in Sweden. Although we were keen to recruit greater numbers of business students, they were underrepresented in the final cohorts. All participants from Tribhuvan University were Nepali nationals. In contrast, the participants from Stockholm University represented a wide range of nationalities. Of the 60 participants from Stockholm, 25 nationalities were represented: 19 Swedish; three mixed nationality Swedish (Swedish – Iraqi, Swedish – Portuguese, and Swedish – Russian); eight German; five Indian; three Dutch and British; two each French, Italian, Spanish, Pakistani; one each Brazilian, Bulgarian, Canadian, Danish, Georgian, Greek, Polish, Romanian, South African, Turkish, and Ukrainian.

Table 1 Study Participants

Trait	Nepal	Sweden
Age		
Mean	23.5	24.2
SD	2.7	5.1
Gender		
Male	44	26
Female	25	34
Academic pathway		
STEM	44	43
Business	25	17

Table 2 Overview of Data Collection and Analytical Procedures

Stage	Time frame: Sweden	Time frame: Nepal	Research instrument	Research question (RQ)	Data collected	Analytical procedures
1	3 weeks (October 2017)	3 weeks (March 2018)	Reading log	RQ1	Sweden: $N = 9$; Nepal: $N = 19$. Name of material read; genre; reading purpose; language; number of pages	Descriptive; Summarizing of texts, genres and languages across research locations
2	Undertaken sequentially (November 2017)	Undertaken sequentially (April 2018)	Reading practices questionnaire (Section 1 may be viewed in Appendix A)	RQ1	Sweden = 60; Nepal = 69. Likert-type items ($n = 41$) on reading practices, linguistic information and attitude to studying in English	Descriptive analysis of Likert-type items; hypothesis testing of group differences
			Administration of the reading section of the TOEFL iBT test	RQ2	Forty-one multiple-choice questions based around three passages	Descriptive; classical item analysis; hypothesis testing of group differences
			TOEFL suitability questionnaire	RQ2	Ten Likert-type items asking students to reflect on their experience of taking the TOEFL iBT reading test section, and questions on previous test experience	Descriptive analysis of items
3			Posttest interviews (Appendix B)	RQ1/2		Manual coding of verbal data; grounded analysis of emergent themes

Data Collection Procedure and Instruments

The two research questions were addressed through the following data collection procedures. The data collection procedure was replicated across two locations to maximize consistency and ensure comparability of findings. The same research instruments (presented in English) were used in both locations, with very minor changes to account for location (e.g., name of university or country). All Likert-type questions remained consistent across EMI contexts. For each location, the following data collection procedure was followed (see Table 2).

Students were invited to participate in the research with the assistance of local researchers who advertised the research on-site. The sample was therefore self-selecting as the research team did not possess the authority or means to conduct stratified random sampling. Prospective participants had some flexibility in how they participated in the research. At a minimum, participants could undertake Stage 2 of the research (questionnaires and reading test). They were offered the opportunity to also complete the reading log (Stage 1) and participate in TOEFL suitability interviews (Stage 3). Participants were asked to sign consent forms for each stage of the research that they agreed to participate in and were informed that they could withdraw at any time and that their data would not be used if they decided to withdraw from the

research. They were offered incentives in the form of vouchers paid for each part of the research in which they participated. They were paid equally in both research locations. Participants were paid the equivalent of GBP£20 for participation in the test and surveys and an additional GBP£10 each for the optional interview and reading log. In both contexts, we were able to secure support from on-site academics who also assisted in recruiting student participants. It was necessary to negotiate access with class teachers in both contexts in order to advertise directly to students, and we, therefore, relied upon these "local champions" to assist us in recruiting the desired numbers.

Students who agreed to take part in Stage 1 (Sweden, $N = 9$; Nepal, $N = 19$) were asked to complete a reading log over a period of 3 weeks. Due to difficulties in negotiating access to research locations, a 3-week period was the maximum time frame we were able to schedule prior to commencing the research on-site. A 3-week period would not capture variation in reading load/type across the academic term, but this data would capture a snapshot of students' reading activities during this period without placing undue pressure on them that a more extensive reading log would entail. A reading log, or learning diary, "is a first-person account of a language learning [and engagement] … documented through regular candid entries … and then analysed for recurring patterns" (Bailey, 1990, p. 215). Reading logs act as a personal record of materials read by students in these contexts. The logs provide an indication of reading load, how much time students devote to reading in English compared to other languages, and the challenges they face. Leung (2002) noted that readers may avoid reading longer texts as it is difficult and time-consuming, although extensive reading may become part of daily activity if readers are given a chance to develop their intrinsic motivation over the 3-week period. The logs also assisted the team to understand whether students in these EMI contexts had a positive self-concept as a reader, in which they self-reported a desire and tendency to read and reported enjoyment of or interest in reading, or whether they experienced negative self-concept and expressed a desire and tendency to avoid reading and reported disliking the activity (Sainsbury & Clarkson, 2008).

Originally, the research team planned to use the data from the open-ended questions in the reading logs to inform the questionnaire design. However, only two reading logs were completed and returned prior to the commencement of Stage 2. For this reason, the main reading practices questionnaire was developed from existing examples available in the literature related to domain analysis and reading strategies (e.g., Cohen & Upton, 2006; Khalifa & Weir, 2009; Weir et al., 2012) and adapted to our own research needs. The questionnaire included five sections. The first section (unnumbered) contained open-ended questions for students to self-report language proficiency and use (see Appendix A). Sections 1–4 contained Likert-type items in a five-option format. Section 1 asked respondents to estimate what proportion of their course language, reading, and assessment materials were in English. Options were 0%, 25%, 50%, 75%, and 100%. Section 2 asked students to state their level of agreement to five statements regarding their enjoyment and confidence when engaging in English. Section 3 asked respondents about their reading habits and level of support from course lecturers. This section contained 11 items. Section 4 asked respondents about reading strategies. Response options were consistent across Sections 3 and 4. Options were 1 (*never*), 2 (*rarely*), 3 (*sometimes*), 4 (*frequently*), and 5 (*always*). The focus of the questionnaire was to inform a profile of the domain of reading for academic purposes in both locations via students' reading experiences, strategies, and difficulties in order to investigate the relationship between emergent profiles and the academic reading construct measured in the reading module of the TOEFL iBT.

Once participants had completed the reading practices questionnaire, they completed the reading section of the TOEFL iBT test and then completed a TOEFL suitability questionnaire immediately afterward. The suitability questionnaire contained 10 Likert-type items, with a 5-point scale of agreement (1 = *strongly disagree*, 2 = *disagree*, 3 = *uncertain*, 4 = *agree*, 5 = *strongly agree*) relating to the participants' general impression of the reading section of the TOEFL iBT test, its suitability for making claims about language proficiency in their contexts, and whether they felt that the test provided an accurate picture of their own language proficiency. These items were also designed to stimulate discussion in the TOEFL suitability interview. An Internet-based version of the test was considered but rejected due to a lack of Internet connectivity in Nepal, which would have severely hindered data collection. Therefore, to maintain consistency, a paper-based version was used in both contexts. The research test form contained three passages and 41 items. The reading practices questionnaire, the TOEFL iBT reading test and TOEFL suitability questionnaires were completed successively by all participants. Completion of all three elements took approximately 90 min total. When completing the reading practices questionnaire, participants were asked if they were willing to take part in an interview to discuss the test and their experiences using EMI. We interviewed all those who stated they were willing, asking them to complete a consent form. Where possible, interviews commenced immediately after the test sessions. If no researcher was available, then the interview would take place later

the same day or the following day. Interviews lasted 10 – 20 min depending on how much the participants contributed. Whereas the interviews in Sweden were conducted in English only, in Nepal some were in English and the majority of the interviews were in Nepali due to the participants' choices. The interview protocol can be seen in Appendix A. Protocols for the interviews were developed simultaneously alongside the questionnaires. This simultaneous development helped to ensure that questions in the interview protocols were pertinent to the study and that the reading practices questionnaire acted as a "primer" for participants, as they would have been thinking about their reading practices prior to the interview. Holding interviews immediately after completing the TOEFL iBT Reading test meant that their recollections of the test were fresh.

Data Analysis

Methods of Quantitative Data Analysis

Test results from Nepali and Swedish contexts were compared directly for overall test score and for each purpose for reading (basic comprehension, inferencing, and reading to learn). Mean, standard deviation, standard error of the mean, and significance testing results, including effect sizes, were calculated to estimate any performance differences across EMI contexts. Additionally, classical item analysis including reliability, item facilities (proportion of respondents who answered correctly), point-biserial correlations (item discrimination), and the proportion of respondents who selected each answer to the multiple-choice items (distractor analysis) were conducted. Because classical item analysis indicates that the tests provide reliable measures of English language proficiency in both contexts, further analysis of quantitative questionnaire data and qualitative interview data is used to explore any emergent difference in proficiency across the contexts and what this analysis tells us about EMI as a phenomenon.

Questionnaire data based on Likert-type items is presented descriptively. Each item is presented with the number of responses for each response category with percentages. Means and standard deviations are also calculated for each item. The intention behind the collection of this type of data was to provide evidence of the types of strategies and processes that are deemed to be important by respondents for their studies. Similarities and differences in response patterns are then compared across the two cohorts.

Methods of Qualitative Data Analysis

Interviews were conducted with students in Sweden and Nepal. In Nepal, 14 interviews were conducted with STEM students and seven with business students. In total, the student interviews represented just over 6 h of audio material (averaging just over 17 min each). Interviews in Nepal were mainly conducted in Nepali. Six interviews were conducted in English when a Nepali-speaking interviewer was unavailable. Ideally, all interviews should have been in the same language. However, we would have lost valuable interview data if we were not flexible on this point. An independent, fluent Nepali speaker translated and transcribed the Nepali interviews into English for the purposes of analysis. These transcriptions were then passed to a second independent Nepali speaker, who compared them to the recordings to ensure the accuracy of the translations. The second translator then discussed her findings with the research team and reported that no discrepancies were found that would lead to ambiguities or misunderstanding in the intention of the interviewees. Instances of any disagreement with the first translator were raised with the research team, which also included a Nepali speaker (second author, Prithvi N. Shrestha), to ensure that no significant ambiguities occurred. In Sweden, 23 interviews were held with students, accounting for 7.5 h of audio (average 19.5 min each). All interviews were conducted in English. These interviews were transcribed independently; when transcriptionists noted moments in the recordings that were difficult to transcribe, researchers checked the transcriptions against the original recordings.

All completed transcriptions were uploaded to QSR NVivo 10 qualitative coding software for analysis. Each of the researchers independently coded three interviews to devise a coding scheme. Then, the research team held a series of three meetings to compare similarities and differences in the codes that had been created and to develop a revised coding scheme. Within NVivo, individual categories are referred to as *nodes*, with subcategories referred to as *child nodes*. Before the second meeting, the scheme was applied to three other interview transcriptions. In the second meeting, the team compared how it had applied the coding scheme to the three transcriptions. In the third meeting, the team made final modifications to the scheme and agreed a procedure to apply the coding scheme to all remaining transcripts. Each

researcher then applied the revised scheme to five further interviews and retrospectively applied the coding scheme to the original coded interviews. The codes were applied to the verbalizations using the following conventions which the research team agreed upon in the third meeting:

- Each part of text (word/sentence level) was coded once only, although utterances may be differentiated if interviewees made unsolicited comments covered by a different node.
- If interviewees made multiple simultaneous comments that related to a single point, these were coded as one utterance.
- If interviewees returned to a point they made earlier, this verbalization was coded independently of the earlier reference to this point. Returning to a point was indicative that the interviewee considered it to be important. Therefore, the increased number of codes for this point would reflect the level of importance. For example, if an interviewee mentioned a specific advantage of studying in English, then made the same point later in the interview, this verbalization would be coded separately.
- Anything related to classroom talk was coded under the category, *seminars*.
- One of the coding categories we used was *diversity*. "Diversity" is understood in terms of international composition of students on course/program, and as such, it encompasses comments related to national and linguistic diversity. In the context of Nepal, this category also included local languages and dialects.
- Within more general nodes, such as *studying in one's own language* or *studying in English*, we created subcategories to record whether the comments were favorable, unfavorable, or neutral in order to determine overall opinion among the interviewees to allow comparisons between Stockholm and Tribhuvan. In NVivo, these subcategories are termed *child nodes*.

Findings and Discussion

Overview of EMI Contexts in Tribhuvan and Stockholm

As part of the questionnaire and interviews, participants were asked to report the linguistic diversity of their study environments to provide context to the research findings. The average number of languages spoken as reported by students from Stockholm University was 2.7 compared to 2.6 from Nepal; that is roughly equivalent.

Nepali students reported knowledge of a range of languages endogenous to Nepal whereas students in Sweden reported knowledge of languages exogenous to Sweden. As English is a commonly used language in education in Nepal, all participants in Nepal reported proficiency in Nepali and English. Additionally, 42 participants reported a working knowledge of Hindi. One participant each reported knowledge of local languages including Tharu, Newari, Maithili, Doteli, and Chaudhary. Three students reported some proficiency with other languages they had previously studied, including Japanese, German, and Italian.

Students from Stockholm University tended to report proficiency in the language of their home country in addition to English. Only two non-Swedish students studying in Stockholm reported proficiency with the Swedish language (one German and one Dutch student). Of the 19 Swedish students, 17 reported proficiency in Swedish and English. One stated proficiency in French and one in Spanish. The average number of languages in Stockholm was raised by the one South African student who reported proficiency in five languages, including English, Xhosa, Setswana, Sesotho, and Afrikaans.

Eight students in Stockholm commented on the diversity of their programs, whereas no students from Tribhuvan commented on diversity. It would appear then that the exogenous linguistic diversity characterizing the Swedish setting is marked (that is, noticed and topicalized), whereas the endogenous linguistic diversity characterizing the Nepali setting is unmarked (that is, seemingly unnoticed and unremarked upon).

Linguistic Background of Participants From Interview Data

Within the interviews, two codes applied to interview data covered linguistic diversity of the participants (see Table 3).

Students at Tribhuvan were far less likely to comment on the diversity of the student body or their language background than students at Stockholm. However, Tribhuvan students were keenly aware of the importance of language background on their higher education, as English-medium education remains a very significant issue in Nepal as a potential source

Table 3 Interview Data for Linguistic and Cultural Background

Theme	Total number of sources	Total number of coded comments	Sweden total coded comments	Nepal total coded comments
Diversity of student body	8	18	18	0
Language background	38	127	57	70

of inequality. It was clear from the interview data with Stockholm students that the cultural and linguistic diversity of the student population was noticed and often viewed positively, as illustrated in the following quotes:

> here's a good mix. Not quite 50/50 [Swedish and overseas students] but we're heading close, I think. We have people from all over the place. Like I've met some Americans, I think someone's come from Australia to study, all sorts of places within Europe. I met a guy from Bosnia the other day who's on my course, Germany, all kinds of different countries. But then there are also quite a lot of Swedes who wanted to do it in English because either they've spent a bit of time living in England, or they have English family, or they just really love speaking in English more than Swedish. (British Bachelor or Science [BSc] student)

> the group that we hang out with like 15 people, we're all, each and every one of us are from different countries. (Greek BSc student)

> and then additionally they also would like different culture perspectives, and different decision-making aspects, as in how do people do in Cameroon, or how do people do in Colombia? How do they act on these current issues like globalisation, like environmental change, like social development? And these I think they added together, and that's why they keep it in English, they offered English from the beginning. And so they really want this international programme for the sake of it, because everyone can learn even more from people that not only study something else, but also have a different personal background. (Dutch BSc student)

Research Question 1: What Are the Reading Practices of Students in the Two EMI Contexts Under Investigation?

Research Question 1 was addressed through reading log data, quantitative questionnaire data, and qualitative interview data. Before discussing the research question in depth, the overall findings of the reading logs, questionnaires, and interview codings are presented.

Findings From Reading Log Data

Table 4 displays an overview of the reading log data obtained from 9 Stockholm University and 10 Tribhuvan University students. It shows the total reading load reported by the participants across the 3-week period and broken down into academic and nonacademic texts and language (in pages and percent). As can be seen, the participants reported that approximately half of their reading over the course of a 3-week period consisted of academic materials, of which the vast majority was in English.

The majority of the reading requirements in Nepal are in English, with a minority of Nepali and some materials in more than one language. This mix of languages occurred as some materials had been translated from Nepali to English, retaining some of the original language in the translations. Table 5 shows reading genres in greater detail.

In interviews, Tribhuvan students reported reading books and articles in English produced by Nepali authors and published in Nepal. However, they also frequently read works produced and published in the United Kingdom and the United States as well as articles from English-language websites. This finding is consistent with reading patterns in Sweden. Stockholm students read a variety of sources in English from authors and publishers both from and outside their academic context. From a reading point of view, the overwhelming majority of texts encountered in both EMI contexts

Table 4 Reading Log Overview: Material Read in Each Language (In Pages and Percent)

Item	Tribhuvan University ($n = 19$)			Stockholm University ($n = 9$)		
Source	Language	Pages	Percent	Language	Pages	Percent
Academic	English	2,008	45.75	English	2,498	50.91
	Nepali	129	2.94	Other	44	0.90
	English/Nepali (mixed)	155	3.53			
Non-academic	English	999	22.77	English	1,520	30.98
	Nepali	1,382	31.48	Other	845	17.22
Total	All	4,390	100		4,907	100

Table 5 Sources Cited by Reading Log Participants

Source	Number of sources cited	Sum of pages	Pages per student
Tribhuvan University ($n = 19$)	171	4,390	231
Article	60	746	39.3
Book/textbook	71	2,818	148.3
Lecture notes	7	163	8.6
Newspaper	4	36	1.9
Reference	8	285	15.0
Thesis	6	307	16.2
Web page	15	35	1.8
Stockholm University ($n = 9$)	139	4,907	545.2
Article	54	777	86.3
Book/textbook	75	3,844	427.1
Lecture notes	3	66	7.3
Reference	5	218	24.2
Web page	2	2	0.2
Grand total	310	9,297	

represented a standard variety of English. McKay and Brown (2015) recognized that English taught across the globe and across nationalities is fairly homogenous in terms of grammar, and the reading log data supports this assertion.

Although the overwhelming majority of academic sources in both contexts were delivered in English, participants in both contexts reported reading some academic sources in languages other than English and expressed an ability to quickly transition between reading academic sources in different languages. Tribhuvan students reported reading a higher proportion of L1 materials for nonacademic reasons than Stockholm students did. Non-academic English language materials read by Tribhuvan students were almost exclusively web content with some newspaper articles. In contrast, Stockholm students read more nonacademic English language materials, which consisted of books, articles, and web resources. The non-English language materials that Stockholm students read covered books, articles, and web pages. Other languages cited by Stockholm students were German, Swedish, Greek, Bulgarian, Georgian, and German. The figures were skewed by one individual who reported reading a novel in their first language over the course of several days for which the reading log was maintained.

Reading log participants reported reading nine different text sources. The overwhelming majority of sources read were textbooks and peer-reviewed academic articles. Due to the large number of STEM participants, reference works were cited that contain key definitions for scientific concepts. Although Tribhuvan students reported a greater number of sources, Stockholm students reported reading a greater number of total pages, suggesting that they read the reported sources in more detail or that the sources they read were on average much longer. Stockholm students averaged just over 35 pages per source, whereas Tribhuvan students averaged just over 25 pages per source. Per student, Stockholm students reported reading more than double the number of pages that Tribhuvan students reported reading within the same time frame. It is possible that this difference in reading load across the two contexts may influence other features of interest, such as the selection of reading strategies. For example, relatively time-consuming strategies, such as asking for translation, might not be feasible under a heavier reading load. Differences in reported reading strategies are investigated in the next section.

Table 6 Questionnaire Section 1 Results for Stockholm University and Tribhuvan University Students

Proportion of English	Context	Options, *N* (%) 0%	25%	50%	75%	100%
1.1 What proportion of the courses on your study program is delivered in English?	Stockholm	1 (1.67)	3 (5.00)	3 (5.00)	4 (6.67)	49 (81.67)
	Tribhuvan	0 (0.00)	2 (2.90)	9 (13.04)	38 (55.07)	20 (28.99)
1.2 What proportion of the reading material is in English?	Stockholm	0 (0.00)	2 (3.33)	1 (1.67)	13 (21.67)	44 (73.33)
	Tribhuvan	0 (0.00)	1 (1.45)	6 (8.70)	17 (24.64)	45 (65.22)
1.3 What proportion of the assessment material (exams, assignments, etc.) is in English?	Stockholm	0 (0.00)	3 (5.00)	4 (6.67)	3 (5.00)	50 (83.33)
	Tribhuvan	0 (0.00)	2 (2.90)	7 (10.14)	11 (15.94)	49 (71.01)

Note. On the individual questionnaires, participants were asked to put a tick (✓) in the column that best describes their level of agreement each statement.

Table 7 Questionnaire Section 2 Results for Stockholm University and Tribhuvan University Students

Studying in English	Context	Level of agreement, *N* (%) Never	Rarely	Sometimes	Frequently	Always
2.1 I enjoy studying in English	Stockholm	0 (0.00)	0 (0.00)	2 (3.33)	22 (36.67)	36 (60.00)
	Tribhuvan	0 (0.00)	0 (0.00)	3 (4.35)	22 (31.88)	44 (63.77)
2.2 I wish the teaching language was my first language	Stockholm	10 (16.95)	14 (23.73)	21 (35.59)	6 (10.17)	8 (13.56)
	Tribhuvan	2 (2.90)	5 (7.25)	19 (27.54)	18 (26.09)	25 (36.23)
2.3 I understand what the lecturer says when they speak in English	Stockholm	0 (0.00)	1 (1.67)	2 (3.33)	17 (28.33)	40 (66.67)
	Tribhuvan	1 (1.45)	2 (2.90)	2 (2.90)	28 (40.58)	36 (52.17)
2.4 I actively participate in classroom discussions in English	Stockholm	0 (0.00)	2 (3.33)	13 (21.67)	16 (26.67)	29 (48.33)
	Tribhuvan	0 (0.00)	3 (4.35)	17 (24.64)	21 (30.43)	28 (40.58)
2.5 I am confident when speaking in English	Stockholm	0 (0.00)	2 (3.33)	11 (18.33)	20 (33.33)	27 (45.00)
	Tribhuvan	1 (1.47)	2 (2.94)	17 (25.00)	31 (45.59)	17 (25.00)

Note. On the individual questionnaires, participants were asked to put a tick (✓) in the column that best describes their level of agreement each statement.

Findings From Reading Practices Questionnaire Data

We received a total of 129 completed questionnaires (60 from Stockholm; 69 from Tribhuvan). This part of the report presents the findings for each section of the questionnaire independently. Raw data is displayed for each question for each university. Each cell shows the number of respondents who selected that option. Percentages are also included in a separate row. Table 6 shows the findings for Section 1, which asked participants to estimate what proportions of their learning occurred through the medium of English. This section was composed of three items.

As shown in Table 6, more than 80% of students based in Stockholm stated that all of their courses were delivered verbally in English. Conversely, students at Tribhuvan reported that their courses were not always delivered in English, with most respondents reporting about 75% of delivery through the medium of English (Item 1.1). This finding is explored in greater detail in the interview data. However, we found greater agreement regarding reading materials, with 65% of Tribhuvan students and 73% of Stockholm students reporting that all of their materials were in English (Item 1.2). However, this left a portion of the students (27%–35%) who actively sought out reading materials not in English. This figure was slightly higher for Tribhuvan students, a finding that is consistent with the reading log data, which also confirmed that Nepali language materials were a consistent feature of their courses. Greater consensus was reached regarding assessment materials, with most students in both locations (greater than 70%) reporting their assessments were undertaken in English (Item 1.3).

Table 7 details Section 2 of the questionnaire. This section asked participants to relay their confidence and satisfaction in using EMI and contained five items.

An overwhelming majority of students in both locations reported that they enjoyed studying in English, with no students in either location disagreeing with the statement, *I enjoy studying in English* (Item 2.1). Similarly, the overwhelming majority of students in both locations reported that they had no difficulty understanding lectures delivered in English and that they were able to actively participate in EMI classes (Items 2.3 and 2.4, respectively). Students in both contexts

Table 8 Questionnaire Section 3 Results for Stockholm University and Tribhuvan University Students

Reading course material in English		Context	Level of agreement, *N* (%)				
			Never	Rarely	Sometimes	Frequently	Always
3.1	The course material is in English	Stockholm	0 (0.00)	0 (0.00)	1 (1.67)	12 (20.00)	47 (78.33)
		Tribhuvan	0 (0.00)	0 (0.00)	4 (5.80)	14 (20.29)	51 (73.91)
3.2	I read the required English course material	Stockholm	0 (0.00)	0 (0.00)	5 (8.33)	18 (30.00)	37 (61.67)
		Tribhuvan	1 (1.45)	1 (1.45)	3 (4.35)	17 (24.64)	47 (68.12)
3.3	I understand the English course material that I read	Stockholm	0 (0.00)	0 (0.00)	1 (1.67)	30 (50.00)	29 (48.33)
		Tribhuvan	0 (0.00)	3 (4.41)	2 (2.94)	38 (55.88)	25 (36.76)
3.4	I wish the course material was in my first language	Stockholm	14 (24.14)	13 (22.41)	24 (41.38)	3 (5.17)	4 (6.90)
		Tribhuvan	4 (5.88)	8 (11.76)	21 (30.88)	14 (20.59)	21 (30.88)
3.5	I read academic texts in English even if they are not part of the course reading list	Stockholm	1 (1.67)	10 (16.67)	15 (25.00)	20 (33.33)	14 (23.33)
		Tribhuvan	0 (0.00)	3 (4.35)	25 (36.23)	26 (37.68)	15 (21.74)
3.6	Lecturers explain why I have to read a certain text	Stockholm	2 (3.33)	8 (13.33)	26 (43.33)	19 (31.67)	5 (8.33)
		Tribhuvan	0 (0.00)	7 (10.29)	15 (22.06)	29 (42.65)	17 (25.00)
3.7	Lecturers provide guidance on appropriate reading strategies for each text	Stockholm	13 (21.67)	24 (40.00)	14 (23.33)	6 (10.00)	3 (5.00)
		Tribhuvan	1 (1.47)	13 (19.12)	19 (27.94)	20 (29.41)	15 (22.06)
3.8	I read popularized material related to my subject in English (e.g., magazines, blogs, newspaper articles, etc.)	Stockholm	1 (1.67)	5 (8.33)	20 (33.33)	21 (35.00)	13 (21.67)
		Tribhuvan	1 (1.47)	2 (2.94)	13 (19.12)	30 (44.12)	22 (32.35)
3.9	I read popularized material related to my subject in my first language (e.g., magazines, blogs, newspaper articles, etc.)	Stockholm	5 (8.33)	13 (21.67)	18 (30.00)	14 (23.33)	10 (16.67)
		Tribhuvan	1 (1.45)	17 (24.64)	18 (26.09)	21 (30.43)	12 (17.39)
3.10	Reading popularized material related to my subject helps me understand course concepts	Stockholm	1 (1.67)	6 (10.00)	22 (36.67)	15 (25.00)	16 (26.67)
		Tribhuvan	0 (0.00)	3 (4.41)	7 (10.29)	23 (33.82)	35 (51.47)
3.11	I prefer reading online to print materials	Stockholm	10 (16.67)	19 (31.67)	17 (28.33)	9 (15.00)	5 (8.33)
		Tribhuvan	2 (2.90)	12 (17.39)	14 (20.29)	20 (28.99)	21 (30.43)

Note. On the individual questionnaires, participants were asked to put a tick (✓) in the column that best describes their level of agreement each statement.

were motivated and read beyond course materials, as shown in the reading log data. However, students at Tribhuvan were more likely than students at Stockholm to report a desire for lectures to be delivered in their first language, reflecting the cognitive challenge of parsing academic content delivered in a second language (Item 2.2). Additionally, students at Tribhuvan were more likely than students at Stockholm to report that they were less confident when speaking in English (Item 2.5).

Section 3 of the questionnaire asked about students' experiences reading in English, their reading habits, and level of support from course lecturers (Table 8). Section 3 comprised 5-point Likert-type items. Respondents reported their levels of agreement with each of 11 affirmative statements.

All students in both contexts confirmed that the reading materials they encountered were in English (Item 3.1), and the majority reported completing the required reading (Item 3.2) and being confident of understanding what they read (Item 3.3). However, a complex picture of participants' attitude toward reading in English emerged in the questionnaire. Despite professing confidence in their English reading ability, a majority of students at Tribhuvan (51.47%) agreed or strongly agreed with the statement, *I wish the course material was in my first language* (Item 3.4). In Stockholm, 41.38% of respondents sometimes felt this way. This perceived lack of confidence in English could be due to the lack of guidance received from course tutors about academic reading and how to manage breakdowns in comprehension. More than half (61.67%) of Stockholm-based students reported that they did not receive sufficient guidance on how to approach English language texts or what to do if they encountered difficulty (Item 3.7). In Tribhuvan, approximately half of the participants also reported feeling that they did not receive sufficient guidance, although they were more likely than the Stockholm students to report that they did receive such guidance overall. As a result, a majority of participants in both contexts sought reading materials related to their subject in their own language as they believed that this was beneficial for their learning (Items 3.9 and 3.10). Despite this lack of guidance, participants in both contexts were sufficiently motivated to seek out and read other English language sources that were not part of their reading list (Item 3.8). In terms of platform,

Table 9 Questionnaire Section 4 Results for Stockholm University and Tribhuvan University Students

			Level of agreement, *N* (%)				
Reading course material in English		Context	Never	Rarely	Sometimes	Frequently	Always
4.1	Read the material quickly to get an overview	Stockholm	0 (0.00)	9 (15.00)	17 (28.33)	25 (41.67)	9 (15.00)
		Tribhuvan	3 (4.35)	5 (7.25)	23 (33.33)	23 (33.33)	15 (21.74)
4.2	Search for specific pieces of information	Stockholm	0 (0.00)	0 (0.00)	17 (28.33)	29 (48.33)	14 (23.33)
		Tribhuvan	0 (0.00)	2 (2.90)	21 (30.43)	20 (28.99)	26 (37.68)
4.3	Identify the main points the author makes	Stockholm	0 (0.00)	0 (0.00)	14 (23.73)	21 (35.59)	24 (40.68)
		Tribhuvan	1 (1.47)	1 (1.47)	9 (13.24)	28 (41.18)	29 (42.65)
4.4	Identify supporting details/evidence for each main point	Stockholm	0 (0.00)	3 (5.00)	15 (25.00)	27 (45.00)	15 (25.00)
		Tribhuvan	0 (0.00)	5 (7.25)	17 (24.64)	25 (36.23)	22 (31.88)
4.5	Link ideas together to understand the whole argument	Stockholm	0 (0.00)	2 (3.33)	12 (20.00)	31 (51.67)	15 (25.00)
		Tribhuvan	0 (0.00)	0 (0.00)	13 (18.84)	30 (43.48)	26 (37.68)
4.6	Link information I read to my existing knowledge	Stockholm	1 (1.67)	1 (1.67)	9 (15.00)	24 (40.00)	25 (41.67)
		Tribhuvan	0 (0.00)	0 (0.00)	10 (14.71)	26 (38.24)	32 (47.06)
4.7	Form questions in my head related to the material	Stockholm	0 (0.00)	8 (13.56)	17 (28.81)	22 (37.29)	12 (20.34)
		Tribhuvan	0 (0.00)	2 (2.94)	25 (36.76)	16 (23.53)	25 (36.76)
4.8	Read a section multiple times to ensure I understand it	Stockholm	0 (0.00)	2 (3.39)	10 (16.95)	32 (54.24)	15 (25.42)
		Tribhuvan	1 (1.45)	5 (7.25)	7 (10.14)	24 (34.78)	32 (46.38)
4.9	Use a dictionary, thesaurus or glossary to check the meaning of unfamiliar words	Stockholm	2 (3.33)	12 (20.00)	11 (18.33)	15 (25.00)	20 (33.33)
		Tribhuvan	1 (1.47)	1 (1.47)	10 (14.71)	14 (20.59)	42 (61.76)
4.10	Consult social media or video sites such as YouTube or TED to help me understand difficult concepts or words	Stockholm %	3 (5.00)	16 (26.67)	17 (28.33)	15 (25.00)	9 (15.00)
		Tribhuvan	2 (2.94)	5 (7.35)	25 (36.76)	18 (26.47)	18 (26.47)
4.11	Read words or sections aloud to help understand unfamiliar words and concepts	Stockholm	10 (16.67)	14 (23.33)	20 (33.33)	10 (16.67)	6 (10.00)
		Tribhuvan	5 (7.46)	17 (25.37)	22 (32.84)	12 (17.91)	11 (16.42)
4.12	Ask friends/relatives to help me understand unfamiliar words and concepts	Stockholm	9 (15.00)	21 (35.00)	17 (28.33)	12 (20.00)	1 (1.67)
		Tribhuvan	2 (2.99)	11 (16.42)	27 (40.30)	14 (20.90)	13 (19.40)
4.13	Try to infer the meaning of unfamiliar words from the context	Stockholm	1 (1.67)	5 (8.33)	12 (20.00)	28 (46.67)	14 (23.33)
		Tribhuvan	1 (1.47)	6 (8.82)	27 (39.71)	20 (29.41)	14 (20.59)
4.14	Translate content into my mother tongue	Stockholm	15 (25.86)	16 (27.59)	18 (31.03)	5 (8.62)	4 (6.90)
		Tribhuvan	8 (11.76)	15 (22.06)	18 (26.47)	12 (17.65)	15 (22.06)
4.15	Make notes which paraphrase the main points of the text	Stockholm	1 (1.72)	15 (25.86)	17 (29.31)	13 (22.41)	12 (20.69)
		Tribhuvan	1 (1.49)	6 (8.96)	17 (25.37)	24 (35.82)	19 (28.36)
4.16	Make notes which link main and supporting points	Stockholm	3 (5.00)	8 (13.33)	15 (25.00)	17 (28.33)	17 (28.33)
		Tribhuvan	2 (2.94)	5 (7.35)	17 (25.00)	32 (47.06)	12 (17.65)
4.17	Compare notes made with notes from previous reading	Stockholm	13 (21.67)	12 (20.00)	22 (36.67)	10 (16.67)	3 (5.00)
		Tribhuvan	0 (0.00)	11 (16.18)	28 (41.18)	13 (19.12)	16 (23.53)
4.18	Write short summaries	Stockholm	5 (8.33)	25 (41.67)	11 (18.33)	14 (23.33)	5 (8.33)
		Tribhuvan	2 (2.94)	15 (22.06)	23 (33.82)	15 (22.06)	13 (19.12)
4.19	Reflect on what I have learned	Stockholm	2 (3.39)	5 (8.47)	20 (33.90)	23 (38.98)	9 (15.25)
		Tribhuvan	1 (1.49)	8 (11.94)	18 (26.87)	28 (41.79)	12 (17.91)
4.20	Reflect on whether my goals (purposes) for reading have been met	Stockholm	7 (11.67)	12 (20.00)	24 (40.00)	10 (16.67)	7 (11.67)
		Tribhuvan	1 (1.47)	5 (7.35)	18 (26.47)	17 (25.00)	27 (39.71)

Note. On the individual questionnaires, participants were asked to put a tick (✓) in the column that best describes their level of agreement each statement.

participants from Tribhuvan reported that they generally preferred online materials to print materials (Item 3.11), a finding that coheres with the reading log data as Tribhuvan-based participants were more likely than Stockholm-based participants to report reading these sources during the 3-week log period. In contrast, Stockholm students were more likely than Tribhuvan students to express a preference for print materials.

Section 4 of the questionnaire was also delivered as a series of 5-point Likert type items. It contained 20 items related to reading strategies (Table 9). Conscious strategies included in the instrument related to careful and expeditious reading, monitoring, and developmental and reflective reading practices, all of which have been shown to interact with reading proficiency (Grabe, 2009; Khalifa & Weir, 2009; Koda, 2005; Phakiti, 2003). As with Section 3, the strategies were written as affirmative statements, with participants reporting their level of agreement.

Almost 80% of students in both locations reported reading to understand the main points an author makes (Item 4.3) as well as to identify additional supporting details (Item 4.4). High-level, intertextual reading was also important, with most students from both locations reporting linking ideas together (Item 4.5) and linking their reading to previous works they had read (Item 4.6), although Tribhuvan students rated comparing their reading to previous reading notes more important

Table 10 Interview Data for Students' Attitudes Toward Course and Course Language

NVivo parent and child node	Total number of coded comments	Sweden total coded comments	Nepal total coded comments
Course language			
Assessment	43	21	22
General	87	51	36
Lectures	28	10	18
Materials	61	20	41
Other	16	6	10
Seminars	15	2	13
Importance of English	48	12	36
Impressions of others' English	12	6	6
Language support	25	19	6
Reading			
Requirements	4	4	0
Strategies	150	38	112
Understanding	100	46	54
Studying in English			
Advantages	107	48	59
Challenges or disadvantages	59	39	20
Studying in one's own language			
Advantages	51	17	34
Challenges or disadvantages	13	6	7

than Stockholm students did (Item 4.17). More than 20% of Stockholm students reported never doing this. Inferential reading was also important as most students in both locations reported that they needed to try to infer the meaning of unknown lexis from context (Item 4.13). Developmental strategies were very important, with most students reporting the use of dictionaries, peers, or social media to share meaning of difficult lexis (Items 4.9, 4.10, and 4.12). However, these strategies were more popular with Tribhuvan students; compared to Stockholm students, nearly twice as many Tribhuvan students reported that they always used dictionaries to look up unfamiliar words. Reflective strategies (Items 4.17, 4.18, 4.19, and 4.20) were reported by participants in both contexts but were universally more widely reported by participants in Tribhuvan than in Stockholm. The use of these time-consuming strategies may link to the reported reading loads in the reading logs. Having read one source, students in Stockholm were more likely to move to another source, whereas at Tribhuvan, participants were more likely to spend time reviewing their reading and checking their understanding with peers and other means of support. High levels of careful local reading were also reported as students tried to interpret what they had read; more than 80% in both locations reported forming questions in their head (Item 4.7) and reading sections multiple times to ensure comprehension (Item 4.8).

The data reveals that although both cohorts employed reading strategies to assist their understanding of English language materials, the strategies were qualitatively different. Stockholm participants were more likely than Tribhuvan students to try and infer the meaning of unknown lexis from context and refer to a dictionary if this strategy was unsuccessful. Conversely, Tribhuvan participants made use of planning strategies that Stockholm participants did. Additionally, they cited their tutors as a source of guidance regarding how to approach individual texts. Although both sets of students reported feeling confident in English, the data revealed some ambivalence by the Nepali students, who were more likely than Swedish students to express a desire for course materials to be presented in their own language.

Findings From Interview Data

Table 10 displays the relevant coding schema for the interviews in Stockholm and Tribhuvan Universities for Research Question 1.

Topics covered during the interviews included the language of the materials used during participants' courses (reading materials and assessment), how interview participants felt about studying in English versus studying in their own language, and specific comments related to reading and reading strategies.

Interview data confirmed the questionnaire data that across EMI contexts all materials presented in reading lists on the curriculum were English language only. Likewise, all assessments, whether exams or essays, were also required to be

written in English. When asked about language preferences, comments made by Nepali students provided an explanation for the finding that Nepali students stated they would prefer more materials in Nepali. Before starting university, students in Nepal attend either Nepali-medium or English-medium schools from an early age. Students who attended Nepali-medium schools perceived themselves to be at a disadvantage compared to those who attended English-medium schools.

> I finished my high school education in Nepali, then right now a Master's in English medium. That's why I'm facing lots of problems, more problems, I can't understand. Maybe one article while I'm reading, two or more times I need to read to understand what is there. (Nepali Master of Science [MSc] student)

As a result, students in Nepal were likely to express a preference for studying in Nepali. Although studying in English was cited as important due to the international nature of the language and the scientific names of particular plants, objects, or concepts, some Nepali students preferred to study in Nepali due to the ease of understanding materials.

> The text, or paragraphs are written in English, so we read it one time, two time, three times but if it is Nepali then one time is enough to understand the concept. (Nepali MSc student)

When asked about reading strategies or approaching difficulties in reading in a second language, interviewees reported a mixture of top-down management strategies and bottom-up word-recognition strategies. Top-down strategies included goal setting, identifying main and supporting points, and information management through creating flow charts or highlighting as shown by these quotes respectively:

> When it's a whole passage that I don't understand then I will re-read it very slowly, and I try to draw in my mind some kind of flow chart of the argument that is presented in the text, so that it becomes more logical to me." (German BSc student)

> I usually try to re-read the whole article, to read it again one more. Because I mean apparently from the first time you cannot understand everything. So, you have to read it one more time, maybe highlight a few important points. I usually do this, I highlight. And then I try to make a structure of the article. (Greek BSc student)

Interviewees also reported idiosyncratic techniques, dependent on their personal learning preferences and their background knowledge. One student reported a preference for form-focused learning strategies such as rehearsing new lexis by reading aloud:

> I just read it aloud again and again until it sort of goes into my brain. Because I'm very much an audio learner, so I just say it again and again until it starts to make sense … [then] oh I recognise that bit, and then that bit relates to that bit of the text. Like where have I see that before in the text? Oh, it's in this paragraph. It's just trying to make little links between things as well. (British BSc student)

The overwhelming majority of language learning strategies reported by participants were bottom-up, lexis-focused strategies. Interviewees reported attempting to identify the meaning of unknown words from context, often by rereading portions of text. Other strategies for comprehension included consulting classmates, dictionaries, indexes, online videos, or Google translate. For retaining the content, students reported rereading the text a number of times, memorizing passages, and taking notes. Nepali students also favored interactive approaches such as asking classmates or teachers. These conversations would take place in Nepali, with interlocutors translating difficult lexis into Nepali.

> If the teacher is there, we can even request our teacher to just translate, or to make us understand in Nepali, we can do that. And if the teacher makes us understand in Nepali, if we don't understand it, it's like that. Or we can also ask friends if we are not that much good with the teachers, we can ask friends. And so, they make us understand in Nepali also. (Nepali MSc student)

Nepali students did not cite top-down text management strategies when reading challenging texts. One interviewee explicitly rejected the idea of using comprehension strategies to manage difficult texts:

> Sometimes even I get difficulties to understand. So, in that case we request teacher, and teacher, they deliver the same thing in Nepali … we don't have to use such strategy actually. (Nepali MSc student)

Some Nepali students attempted to guess meaning from context, but this was an infrequently cited strategy, as most students felt peer support was the principal means of dealing with difficult texts:

> Yeah, I use different strategies. Like there's someone better than me … I'll even ask them the meaning. Sometime if no one's there, and I feel lazy to look for the meaning, then I just guess also. So, it's depends. (Nepali MSc student)

However, in STEM disciplines, direct translation of specific terminology may not be possible. Students then have to learn the vocabulary ("It's difficult to get the direct translation of scientific words, so I will have to understand from English meaning only," Nepali MSc student), although students may refer to staff and classmates to explain the terms in Nepali.

Emergent differences are likely due to the composition of the two cohorts. The Stockholm sample was composed of a mixture of international students and Swedish students electing EMI pathways. This cohort may therefore be self-selecting for English proficiency, with those students who had studied English to a high level more likely to choose EMI pathways. Additionally, they could not necessarily rely upon classmates to be able to translate difficult terminology and could not ask staff to translate directly during class time. In contrast, Nepali students, who did not have much choice regarding the language of instruction as the majority of academic pathways, especially STEM and business, are delivered through EMI, could all converse in Nepali and were far more likely than Stockholm students to rely on classmates or staff for language support.

Across both contexts, test takers were oriented not only to English but also to the other languages in their multilingual repertoires. Therefore, although English is available, it is not necessarily chosen as the only language appropriate to a particular interaction (spoken or written). As a result, students in EMI contexts read from a diverse range of sources, including academic books, journal articles, reference works, online sources, and lecture notes. Students from both contexts reported reading in English for recreational purposes and reading around their topic by consulting English sources. Consistent with these multilingual environments, students exhibit *parallel language use* (simultaneous use of more than one language in one or more domains; Salö & Josephson, 2013, as cited in Källkvist & Hult, 2016) for both recreational and academic reading. This parallel language use was reported by participants as a common real-world practice within both EMI contexts and a pragmatic response to the availability of sources in multiple languages (Hultgren, 2014). Participants did not report concerns that this practice diminished or hindered their learning or ability to teach in English.

English language proficiency is crucial to the success of students' studies in EMI contexts. However, comparability of the relative levels of proficiency across contexts is made challenging due to a lack of specific entry tests. Participants' experiences of taking English language tests vary enormously. In both contexts, English tests taken as part of secondary schooling serve as sufficient evidence of proficiency. Students in Sweden present evidence of English language proficiency in the form of test scores from their home country. In Nepal, English language qualifications used for entry to higher education are those obtained during high school. In Sweden, some participants held language test qualifications but did not report needing them to gain admission to EMI courses in Stockholm.

The rapid expansion of EMI (Airey et al., 2015) represents a challenge for students in both contexts. Students are faced with a sudden demand of learning through a second (or third) language in stark contrast to pre-tertiary levels where students are rarely exposed to EMI. In Nepal, this problem is exacerbated by social stratification as pre-tertiary EMI is limited to affluent families (Phyak, 2016). Specifically, students reported a lack of time to engage with all of the academic materials they would like to read. Inadequate time is compounded by the state of participants' subject knowledge, especially for undergraduate students. The subject knowledge of postgraduate students meant that they were more confident than undergraduates in dealing with English language materials in their subject area. However, postgraduate students in Nepal specifically expressed difficulty in reading in English in areas with which they were unfamiliar. This perception also manifested in TOEFL suitability questionnaire findings in which Nepali students in particular expressed the belief that text topic would have a significant impact on test scores. However, empirical evidence shows that knowledge of a topic because of major of study does not unfairly impact test performance (Liu, 2011). So although the students might think that knowing about the topic of the reading passage will help them perform well on the test (which makes sense, as topical knowledge is important for language comprehension), nevertheless, the test is designed to minimize the unfair impact of knowledge about the topic of the passage.

Table 11 Score Comparisons Across EMI Contexts

Reading purpose	EMI context	*N*	Max score	Mean	*SD*	SEM
Basic comprehension	Nepal	69	29	14.62	5.49	.66
	Sweden	60		24.42	3.26	.42
Inferencing	Nepal	69	9	3.20	1.89	.23
	Sweden	60		6.58	1.85	.24
Reading to learn	Nepal	69	7	2.58	1.84	.22
	Sweden	60		4.33	1.69	.22
Total	Nepal	69	45	20.41	8.17	.98
	Sweden	60		35.83	5.64	.73

The study findings reveal that assistance currently provided by Nepali tutors with English language materials mostly consists of translation. Students either ask tutors to translate during lectures or look up difficult terminology during private study. Little evidence emerged of Nepali tutors providing instruction in reading strategies. As Nepali students cited reliance on tutors, the lack of qualified teachers who can teach in English is a key concern. The growing demand for more EMI schools in the country (Ministry of Education, 2016) means that a new focus on effective reading strategies in teacher training is essential to assisting students to keep up with set readings as they progress through their courses and are presented with more complex topics. Nepali tutors would therefore benefit from targeted training in reading strategies, which would assist students to develop the skills needed to understand English language materials and reduce the need for translation. Conversely, students in Stockholm seemed more confident with independent reading strategies. Students in Stockholm may instead benefit from academic skills training, which should include explicit reading strategies training, such as global reading to obtain gist, inferring the meaning of unfamiliar lexis, and building an overall impression of a text.

Research Question 2: Is the Reading Section of the TOEFL iBT Suitable as an Entry Test for Different EMI Program Courses?

Research Question 2 was addressed through test data, TOEFL iBT suitability questionnaires, and interview questions targeting the suitability of the TOEFL iBT. For the purposes of the study, the primary functions of the reading test were to elicit students' opinions regarding face-validity of the test and document reading proficiency levels within and between groups. Firstly, we briefly present the test data to determine what claims the test results allow us to make in terms of the three reading purposes (basic comprehension, inferencing, and reading to learn). Secondly, we go on to examine TOEFL iBT suitability questionnaire and interview data to examine participants' perceptions of the test.

Comparing English Language Proficiency Across the Two Contexts

We obtained 69 complete test records from Nepal and 60 from Sweden ($n = 129$). The test is divided into three sections, each containing one text. The first text has 14 questions, the second has 13, and the third has 14 (see Sections 3.1 and 3.3 for more information about the test used in this research). Test results from Nepali and Swedish contexts were compared directly for overall test score and for each reading purpose (basic comprehension, inferencing, and reading to learn). Basic comprehension and inferencing items are dichotomous (correct/incorrect). There are three reading-to-learn items: two are worth 2 points and the third is worth 3 points. Therefore, the maximum score that could be obtained in the test is 45 points (29 for basic comprehension, 9 for inferencing, and 7 for reading to learn). Descriptive statistics are outlined in Table 11.

Overall reliability for the test as measured by Cronbach's alpha was strong. The data recorded a reliability coefficient of .93 (.86 for Tribhuvan University students and .80 for Stockholm University students). Prior to comparison, Levene's test (Levene, 1960) was performed to determine whether sample scores from the two cohorts contained equal variance. The outcome ($f = 6.40$, $p < .05$) indicated unequal variance, meaning nonparamatric tests would be required to compare findings across the two contexts. This result is visible in the data by the different standard deviations of the two groups. The spread of scores for Tribhuvan-based students was almost 3 points greater than that of Stockholm-based students (Table 11). Therefore, we used the Mann – Whitney U test to compare scores across the two contexts for each of the three

Table 12 TOEFL Suitability Questionnaire Results

Reading course material in English		Context	Level of agreement, *N* (%)				
			Strongly disagree	Disagree	Uncertain	Agree	Strongly agree
5.1	The different types of questions in the TOEFL reflect the English reading skills needed for my course	Stockholm	0 (0.00)	0 (0.00)	15 (25.00)	36 (60.00)	9 (15.00)
		Tribhuvan	0 (0.00)	0 (0.00)	3 (4.55)	38 (57.58)	25 (37.88)
5.2	A university can learn a lot about a person's English reading ability from their TOEFL performance	Stockholm	0 (0.00)	7 (11.67)	25 (41.67)	23 (38.33)	5 (8.33)
		Tribhuvan	0 (0.00)	4 (5.97)	9 (13.43)	34 (50.75)	20 (29.85)
5.3	The TOEFL test is an accurate reflection of my English reading ability	Stockholm	0 (0.00)	8 (13.33)	22 (36.67)	28 (46.67)	2 (3.33)
		Tribhuvan	0 (0.00)	6 (8.82)	15 (22.06)	35 (51.47)	12 (17.65)
5.4	The TOEFL reading texts were difficult to understand	Stockholm	5 (8.33)	39 (65.00)	8 (13.33)	8 (13.33)	0 (0.00)
		Tribhuvan	1 (1.47)	19 (27.94)	26 (38.24)	15 (22.06)	7 (10.29)
5.5	People who perform well in the TOEFL reading test will do well in university	Stockholm	8 (13.56)	12 (20.34)	23 (38.98)	13 (22.03)	3 (5.08)
		Tribhuvan	2 (2.99)	12 (17.91)	27 (40.30)	18 (26.87)	8 (11.94)
5.6	A bad performance in the test shows that your reading ability is limited	Stockholm	2 (3.33)	19 (31.67)	22 (36.67)	14 (23.33)	3 (5.00)
		Tribhuvan	2 (2.90)	22 (31.88)	13 (18.84)	25 (36.23)	4 (5.80)
5.7	The TOEFL examines English reading skills I will never use	Stockholm	14 (23.33)	36 (60.00)	6 (10.00)	3 (5.00)	1 (1.67)
		Tribhuvan	9 (13.64)	25 (37.88)	20 (30.30)	11 (16.67)	1 (1.52)
5.8	Performance in the test depends on text topic	Stockholm	0 (0.00)	5 (8.33)	7 (11.67)	29 (48.33)	19 (31.67)
		Tribhuvan	1 (1.49)	4 (5.97)	5 (7.46)	39 (58.21)	18 (26.87)
5.9	I feel that my English reading ability is better than my performance will indicate	Stockholm	2 (3.33)	5 (8.33)	33 (55.00)	19 (31.67)	1 (1.67)
		Tribhuvan	0 (0.00)	3 (4.41)	25 (36.76)	29 (42.65)	11 (16.18)
5.10	I am confident I performed well in the test	Stockholm	0 (0.00)	2 (3.33)	25 (41.67)	28 (46.67)	5 (8.33)
		Tribhuvan	0 (0.00)	2 (2.94)	15 (22.06)	40 (58.82)	11 (16.18)

Note. On the individual questionnaires, participants were asked to put a tick (✓) in the column that best describes their level of agreement each statement.

reading purposes and overall scores and computed the effect size of findings using Cohen's *d*. Overall, median score differences were statistically significantly different. Stockholm University students outperformed Tribhuvan University students in the test ($U = 2{,}690.00$, $p < .01$, $d = 0.53$). This difference is repeated for each of the median scores for each of the three reading purposes in the test; basic comprehension ($U = 2{,}686.00$, $p < .01$, $d = 2.17$), inferencing ($U = 2{,}877.00$, $p < .01$, $d = 0.71$), and reading to learn ($U = 3{,}207.50$, $p < .01$, $d = 1.07$). All tests were significant at the Bonferroni-adjusted *p*-value of .01. Although the test data suggest that students in the Swedish context appear to possess a higher level of reading proficiency than students based in Nepal as measured by the reading section of the TOEFL iBT test, we found significant contextual differences between these two EMI universities in addition to contextual differences within the Nepali cohort. Data responses from the reading section of the TOEFL iBT test indicated that that test identified clear differences in reading language proficiency in the two contexts. This finding held for each of the three reading purposes in the TOEFL iBT Reading test (basic comprehension, inferencing, and reading to learn). Additionally, the high standard deviations suggest that the test is sensitive to significant individual variation within contexts, a potential advantage for the test to have a diagnostic function. The differences we observed in reading proficiency in the Nepalese and Swedish context may have implications for the type of remedial assistance that is required in each between the two contexts, rather than employing the test as an explicit gatekeeping device.

Comparing Students' Perceptions of the Reading Section of the TOEFL iBT Test in the Two EMI Contexts

The TOEFL suitability questionnaire was composed of 10 Likert-type items with a five-options scale of agreement (1 = *strongly disagree*, 2 = *disagree*, 3 = *uncertain*, 4 = *agree*, 5 = *strongly agree*). Table 12 shows the results for the section of the questionnaire that compared perceptions of the reading section of the TOEFL iBT test in the two contexts. It shows the extent to which students in each context agreed with a number of statements about the test.

Participants from both contexts broadly expressed positivity regarding the relationship between the types of questions in the reading section of the TOEFL iBT test and the academic English reading skills required for their academic programs (Item 5.1). Tribhuvan-based students were more likely to express positivity regarding the alignment between reading skills represented by the test and those required for their courses relative to Stockholm-based students, of whom one quarter expressed some uncertainty about whether the test reflected the skills required for their course (Item 5.1). This

Table 13 Impressions of the Reading Section of the TOEFL iBT Test: Coded Comments by EMI Context

NVivo parent and child nodes	Total number of coded comments	Sweden total coded comments	Nepal total coded comments
Impressions of taking the TOEFL test			
Mixed	59	26	33
Negative	74	39	35
Neutral	74	58	16
Positive	76	43	33
Suitability of TOEFL	78	29	49

uncertainty increased when asked about the use of TOEFL iBT to make claims about reading ability (Item 5.3) in both contexts, although only 14 participants overall expressed negativity toward the use of reading section of the TOEFL iBT test to make claims about English language reading proficiency. The majority of participants from both contexts agreed or strongly agreed that the TOEFL iBT test is useful for making claims regarding reading proficiency (Items 5.2 and 5.3), although they felt that a bad performance may be due to factors other than reading proficiency (Item 5.6).

The cohort was split regarding the predictive validity of the test as more than one third of the participants expressed uncertainty that students who did well in the TOEFL iBT Reading test would go on to do well at university (Item 5.5). One factor they felt may impact test performance was text topic, as more than 80% of the cohort across both contexts believed this influenced performance (Item 5.8; see, however, evidence from Liu (2011) that despite this perception, this is not empirically corroborated in actual test results to date). However, participants clearly felt that some aspects of the reading section of the TOEFL iBT test were relevant to their reading experience as the overwhelming majority of participants disagreed with the statement that they would never use reading skills embedded in the test (Item 5.7). In contrast to their high scores, participants from Stockholm were generally uncertain about their performance in the test, with 55% of participants stating that their ability was better than their performance indicated (Item 5.10). The difference in test scores between the two contexts is reflected in the difference in perception of TOEFL iBT text complexity. Tribhuvan students were more likely than Swedish students to state that the texts were difficult to understand (Item 5.4). Participants from both contexts were confident that they performed well (Item 5.10), although they also felt that their reading ability could be better than that reported by the test (Item 5.9).

Interviews with participants further examined their attitudes toward the reading section of the TOEFL iBT test. Interviews were conducted immediately after they had taken the test so that their perspectives of the text and tasks remained fresh. This verbal data from participants provided greater detail regarding participants' opinions. Comments were coded based on whether the statements were perceived to be positive, negative, neutral, or mixed. If participants made comments regarding the suitability of the reading section of the TOEFL iBT test for their specific academic pathway, these were coded separately. Table 13 summarizes the comments made by participants regarding their perceptions of its suitability.

Students from both countries offered nuanced comments regarding their impressions of the reading section of the TOEFL iBT test and its suitability for making claims about reading proficiency within their EMI contexts. For Stockholm-based students, more than half of the comments were coded as *neutral*. After examining comments more closely, they were further categorized into cognitive, context, and consequential validity. Comments on validity included coded verbalizations from both the impressions and suitability nodes.

Students made a large number of comments relating to the context validity of the reading section of the TOEFL iBT test. The overwhelming majority of these comments related to the suitability of passage topic and the perceived impact of content knowledge on test performance. Nearly all comments coded as *negative* related to the appropriacy of text topics. Students' perceptions were influenced by their academic pathway. As described in the Methodology section, the reading test contained three passages: one related to psychology, one to biology, and one to paleontology. Students undertaking science courses were satisfied with the appropriacy of the latter two texts but perceived their performance may have been hindered by a lack of familiarity with the terminology related to psychology. This perception was common to both Stockholm and Tribhuvan students.

> The second and third passages were related to fish and biology, fossils, I understood it very well, my performance was good on those but first one was difficult to understand. (Nepali BSc student, Nepal)

> But if it is not of the field of your study, then I think it's not good to evaluate someone by putting some questions which is not of interest of them. I found comfortable to answer the question which were in Science, but I found it extremely difficult to write the answers or tick the sentences for the another one. (Nepali BSc student, Nepal)

> The one about the fossils and osmosis I think I understood most or all of it. The first one I probably mostly forgotten when I get home. (Swedish MSc student, Sweden)

> It mostly depends on the topic of the text. For me it was because I did geography and biology text was amazing but then I found about art and about paintings. And I'm really bad in that, so I could hardly understand. (Georgian MSc student, Sweden)

Despite the lack of empirical support that text topic influences test performance (Liu, 2011), participants' perceptions of the suitability of text topic is an important issue of *face validity*. Students perceived the relevance of passage content as an issue of test fairness as the content of the test should reflect the kinds of texts they engage with in their academic program. Research shows that using texts unrelated to students' field of study does not necessarily impact test performance in a negative way, as TOEFL iBT Reading tasks are purposefully designed to not be dependent on the meaning of complex, subject-specific terminology (Liu, 2011). However, the study participants were not familiar with the relevant research findings. However, one participant from Stockholm did express skepticism at the proposition of subject-specific language tests and expressed an understanding that a test that had too close an emphasis on subject matter would no longer be a test of English reading proficiency:

> But I think it's good to have that standardised medium, to have everyone on the same bench level, so when they come into the course you know ah, this person is like, they've got their English, they'll be fine keeping up, or this person hasn't quite made the marks therefore they're going to struggle. But it's just is it necessarily relevant to the course is the thing. It would be much easier of it was, like the people taking business won't have studied biology. The people taking biology won't have studied business. It's about having something that people could relate to … but it starts to feel like it's becoming more of a subject specific assessment after a while. I mean when it comes to things like oh what's this word closely related to? Then you can tell that's more of a test of your English knowledge. (English BSc student)

The last sentence in this quote also suggests this participant also recognized the utility of the vocabulary task type, as this is clearly a question designed to assess language knowledge rather than content knowledge.

Comments were coded as *cognitive validity* when they referred to the mental activities undertaken in the action of reading. Cognitive validity includes the processing of text and the higher order strategic decisions readers make about how to engage with specific texts. Students' comments covered both the kinds of processes undertaken to complete specific tasks, the kinds of strategies they employed, and the relevance of these to their contexts. Students perceived that the reading load imposed by the test tasks was heavy and this was affected by the timed conditions. A lack of understanding of the difficult text meant that students reread texts several times to ensure comprehension.

> This is also quite difficult for me. I need to read more than two or three times, but time is too much limited in that section, I think. I need to need, if I read an article like my course related articles, to understand that article, what is inside there, what is the main objectives, more than two/three times I need to read. After that I was to consult with seniors, Google search, also talk seniors that I will solve the problem. That's why I need to finish in limited time, that's why I'm quite confused. (Nepali MSc student)

The Nepali MSc student perceived the timed test conditions to be dissimilar to the nature of reading in their context. There was also disquiet regarding the elicited cognitive processes in the test-taking environment, with one student stating that tasks elicited word-level comprehension processing with a deficit in higher, text-level processing:

> University requires to understand the whole meaning of texts and this test was more about certain words or certain sentences or smaller parts. At the end there was a question that's more about the overall [text] … so yes I'm not sure if it's covering the whole [text]. (German MSc student)

However, some students disagreed and cited the perception that the reading test encouraged high-level processing and the suitability of the test.

> I get that that's tricky for some people, but I think it was a really good test, because it's important to understand the meaning of a text where it does not tell us directly what it's supposed to mean. Understanding the paragraph. (Swedish MSc student)

> Anybody who does well in a TOEFL test is probably good at English. Actually, all three questions were very well chosen. Whoever designed this test is really good. (Swedish MSc student)

Students also made positive comments, expressing the belief that completing the test would encourage them to develop their English reading skills.

> I think we can improve our English, we can improve our understanding, so we can improve our reading. And it is more beneficial to us I think, it is beneficial for improving our English. (Nepali MSc student)

> Obviously, we find TOEFL test difficult but beyond the difficulty we also learned something, like this certain techniques must be followed for studying. (Nepali BSc student)

Negative comments centered on the appropriacy of making high-stakes decisions on the basis of a short test.

> Because you have, sometimes you have a bad day or something and you can't concentrate, or you have a long day or something, you feel sick. And you have the one test and you have to really, you have to be good to have the chance in the end to go abroad or something. (German MSc student)

Conclusions

This report has explored reading practices in two highly different EMI contexts (Global North and Global South) and the suitability of the reading section of the TOEFL iBT test to assess reading in such contexts. We argue that this report raises important issues for test providers and developers seeking testing solutions specifically related to academic reading in complex and diverse EMI contexts. This research has revealed that academic reading is a fundamental part of academic success in both EMI contexts, although across these contexts, participants employ different academic study skills and reading strategies to adapt to the resources available to them.

Challenges Faced in This Study

Studying reading in two diverse EMI contexts in different parts of the world represented significant logistical challenges. Gaining access to the two institutions required the assistance of local colleagues who were able to advertise the research with colleagues and, most importantly, prospective participants in the study. Negotiating took longer than expected, meaning there was a delay in commencing data collection. This delay meant that our data collection period for the reading logs was severely truncated. Originally, we had planned to devise our questionnaires to reflect the findings from the reading logs. However, it became necessary for us to draw up the questionnaires before any logs were completed. In the case of Sweden, several participants were still completing their logs whilst participating in the main data collection. Obtaining a minimum of 50 participants in the study would not have been possible without the support of local colleagues. We also originally planned to focus the study on undergraduate students. However, it became clear as we progressed that we would not be able to secure the required numbers if we only focused on undergraduates. As a result, we opened data collection to postgraduate students. In Tribhuvan, we were also faced with the prospect of a disproportionate number of male students in the study, and so we had to ask for further female volunteers in an effort to make the sample more representative. Our time in Stockholm was limited to 1 week, which we realized was insufficient time to complete the required data collection. This shortcoming was remedied in the data collection in Tribhuvan, in which we organized 2 weeks to collect the data, reflected in the greater proportion of undergraduate students in this cohort and a larger number of participants, better reflecting our original research aims.

Testing Academic Reading in Diverse EMI Contexts

Interview, questionnaire, and test data indicated that the TOEFL iBT reading test was useful in identifying English reading proficiency of multilingual students in diverse EMI settings. Also, the reading log data provided evidence that participants in both contexts did not read varietal forms of written English. All English language texts encountered in both EMI contexts represented standard written English (McKay & Brown, 2015). However, the overwhelming perception by participants in this study is that text content impacts their test performance and that tests should therefore reflect their academic pathways to ensure that fair decisions are made. This finding was backed up by interview data, which indicated that students believed that text topic impacted their performance in the TOEFL iBT, although empirical research to date has suggested that this is not a factor (Liu, 2011). Given that a validity argument for test use within a specific context must incorporate the attitudes toward the language contained within test tasks (face validity; Taylor, 2006), the introduction of an English language test in EMI contexts must accommodate the perspectives of local stakeholders.

Although both EMI contexts are multilingual, participants from Nepal represent a single nationality. The data revealed that reading and study skills involved the use of translation, Nepali language and L1-based strategies, such as asking peers or tutors to translate, to a much greater degree than were reported in Stockholm. In Stockholm, English was a much more distinct lingua franca, based on the universality of the language and the multinational make-up of the student cohort.

To make valid claims about preparedness to study in higher education within EMI contexts, an English test must represent the kinds of communication that students encounter in their context. An English language test in Nepal would need to reflect the specific kinds of communication found in this specific setting. However, adapting an international test of English to each local context is not feasible. Additionally, requiring entry-level proficiency tests of incoming students in EMI courses may not be possible because of local or national policies oriented toward local or school-level qualifications as well as the diverse educational backgrounds and pathways of incoming students, such as those in Stockholm who presented with a range of qualifications. Despite these issues, this research has demonstrated that the reading section of the TOEFL iBT test can provide useful information about student academic language proficiency, and could be used to inform remedial English language instruction for incoming students. If tutors possessed more information about incoming students' reading proficiency, such information would assist in the development of appropriate instructional responses and materials.

Acknowledgments

This research was funded by ETS under a Committee of Examiners and the *TOEFL*® research grant. ETS does not discount or endorse the methodology, results, implications, or opinions presented by the researcher(s). We wish to offer special thanks to the many colleagues, students, and academic staff who assisted us during the course of this project, in particular Maria Kuteeva and Spela Mežek at Stockholm University and Bharat Babu Shrestha at Tribhuvan University. Without their support, this research would not have been possible. We wish to thank the anonymous referees, Laura Ballard, Cris Ciobanu, Spiros Papageorgiou, and colleagues at ETS for their many useful insights and suggestions in the preparation of this manuscript. We also wish to acknowledge Stephen Bax, who initiated this project and provided significant guidance and input into the project. He is sorely missed by family, friends, and colleagues. We dedicate this work to Professor Stephen Bax, who continues to inspire colleagues and the students he worked with during the course of his career (Professor Stephen Bax 1960 – 2017)

Notes

1 The concepts of "Global South" and "Global North" refer to the socioeconomic and political division of countries in the world. Global South includes low-income countries in Latin America, Asia, Africa, and Oceania while Global North includes developed countries in America, Europe, Australia, and a few countries such as Japan in Asia (Dados & Connell, 2012).

2 More information about the reading questions in the TOEFL iBT is available in a series of videos at https://www.ets.org/toefl/ibt/about/video_library/

References

Airey, J., Lauridsen, K., Räsänen, A., Salö, L., & Schwach, V. (2015). The expansion of English-medium instruction in the Nordic countries: Can top-down university language policies encourage bottom-up disciplinary literacy goals? *Higher Education*, *73*, 561–576. https://doi.org/10.1007/s10734-015-9950-2

Airey, J., & Linder, C. (2008). Bilingual scientific literacy? The use of English in Swedish university science courses. *Nordic Journal of English Studies*, *7*(3), 145–161. https://doi.org/10.35360/njes.105

Bailey, K. M. (1990). The use of diary studies in teacher education programs. In J.C. Richards & D. Nunan (Eds.), *Second language teacher education* (pp. 215–226). Cambridge University Press.

Bolton, K., & Kuteeva, M. (2012). English as an academic language at a Swedish university: Parallel language use and the "threat" of English. *Journal of Multilingual and Multicultural Development*, *33*(5), 429–447. https://doi.org/10.1080/01434632.2012.670241

Cohen, A., & Upton, T. (2006). *Strategies in responding to the new TOEFL Reading tasks* (TOEFL Monograph Series MS-33). ETS.

Dados, N., & Connell, R. (2012). The global south. *Contexts*, *11*(1), 12–13. https://doi.org/10.1177/1536504212436479

Dearden, J. (2014). English as a medium of instruction—A growing global phenomenon. *British Council*. https://doi.org/10.13140/RG.2.2.12079.94888

English First. (2019). *EF English Proficiency Index: A ranking of 100 countries and regions by English skills*. https://www.ef.co.uk/assetscdn/WIBIwq6RdJvcD9bc8RMd/legacy/__/~/media/centralefcom/epi/downloads/full-reports/v9/ef-epi-2019-english.pdf

Enright, M., Grabe, W., Koda, K., Mosenthal, P., Mulcahy-Ernt, P., & Schedl, M. (2000). *TOEFL 2000 reading framework: A working paper* (Research Memorandum RM-00-4). ETS.

ETS. (2020). TOEFL iBT®: *Test and score data summary 2019*. https://www.ets.org/s/toefl/pdf/94227_unlweb.pdf

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. Cambridge University Press. https://doi.org/10.1017/CBO9781139150484

Hultgren, A. K. (2014). Whose parallel lingualism? Overt and covert ideologies in Danish university language policies. *Multilingua*, *1–2*(33), 61–87. https://doi.org/10.1515/multi-2014-0004

Hultgren, A. K., Jensen, C., & Dimova, S. (2015). Introduction: English-medium instruction in European higher education: From the north to the south. In S. Dimova, A. K. Hultgren, & C. Jensen (Eds.), *English in Europe: Vol. 3. English-medium instruction in European higher education* (pp. 1–16). Mouton. https://doi.org/10.1515/9781614515272-002

Jamieson, J. M., Eignor, D., Grabe, W., & Kunnan, A. J. (2008). Frameworks for a new TOEFL. In C. Chappelle, M. K. Enright, & J. M. Jamieson (Eds.), *Building a validity argument for the test of English as a foreign language* (pp. 55–95). Routledge. https://doi.org/10.4324/9780203937891

Källkvist, M., & Hult, F. (2016). Discursive mechanisms and human agency in language policy formation: Negotiating bilingualism and parallel language use at a Swedish university. *International Journal of Bilingual Education and Bilingualism*, *19*(1), 1–17. https://doi.org/10.1080/13670050.2014.956044

Khalifa, H., & Weir, C. J. (2009). *Studies in language testing: Vol 29. Examining reading: Research and practice in assessing second language reading*. Cambridge University Press.

Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach*. Cambridge University Press. https://doi.org/10.1017/CBO9781139524841

Kuteeva, M., & Airey, J. (2013). Disciplinary differences in the use of English in higher education: Reflections on recent policy developments. *Higher Education 67*(5), 533–549. https://doi.org/10.1007/s10734-013-9660-6

Leung, C. Y. (2002). Extensive reading and language learning: A diary study of a beginning learner of Japanese. *Reading in a Foreign Language*, *14*(1), 66–81.

Levene, H. (1960). Robust tests for equality of variances. In I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, & H. B. Mann (Eds.), *Contributions to probability and statistics: Essays in honor of Harold Hotelling* (pp. 278–292). Stanford University Press.

Liu, O. L. (2011). Does major field of study and cultural familiarity affect TOEFL iBT reading performance? A confirmatory approach to differential item functioning. *Applied Measurement in Education*, *24*(3), 235–255. https://doi.org/10.1080/08957347.2011.580645

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

McKay, S. L., & Brown, J. D. (2015). *Teaching and assessing EIL in local contexts around the world*. Routledge. https://doi.org/10.1080/1554480X.2016.1209821

Mežek, S. (2013). *Advanced second-language reading and vocabulary learning in the parallel-language university* (Unpublished doctoral thesis). Stockholm University.

Ministry of Education. (2016). *School sector development plan, 2016/17-2022/23*. Government of Nepal. http://www.moe.gov.np/assets/uploads/files/SSDP_Book_English_Final_July_5,_2017.pdf

Moore, T., Morton, J., & Price, S. (2007). Construct validity in the IELTS Academic Reading test: A comparison of reading requirements in IELTS test items and in university study. *IELTS Research Reports*, *11*(4), 1–86. https://www.ielts.org/-/media/research-reports/ielts_rr_volume11_report4.ashx

Ockey, G. J., Koyama, D., Setoguchi, E., & Sun, A. (2014). The extent to which TOEFL iBT speaking scores are associated with performance on oral language tasks and oral ability components for Japanese university students. *Language Testing*, *32*(1), 39–62. https://doi.org/10.1177/0265532214538014

O'Dwyer, J., Kantarcıoğlu, E., & Thomas, C. (2018). *An investigation of the predictive validity of the TOEFL iBT® test at an English-medium university in Turkey* (Research Report No. TOELF-RR-83). ETS. https://doi.org/10.1002/ets2.12230

Papageorgiou, S., Wu, S., Hsieh, C.-N., Tannenbaum, R. J., & Cheng, M. (2019). *Mapping the TOEFL iBT test scores to China's standard of English language ability: Implications for score interpretation and use* (Research Report No. TOEFL-RR-89). ETS. https://doi.org/10.1002/ets2.12281

Pearlman, M. (2008). Finalizing the test blueprint. In C. A. Chapelle, M. K. Enright, & J. M. Jamieson (Eds), *Building a validity argument for the test of English as a foreign language* (pp. 227–258). Routledge. https://doi.org/10.4324/9780203937891

Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. *Language Testing*, *20*(1), 26–56. https://doi.org/10.1191/0265532203lt243oa

Phyak, P. (2016). Local-global tension in the ideological construction of English language education policy in Nepal. In R. Kirkpatrick (Ed.), *English language education policy in Asia* (pp. 199–217). Springer International Publishing. https://doi.org/10.1007/978-3-319-22464-0_9

Rogier, D. (2012). *The effects of English-medium instruction on language proficiency of students enrolled in higher education in the UAE* (Unpublished doctoral thesis). University of Exeter.

Saarinen, T., & Nikula, T. (2013). Implicit policy, invisible language: Policies and practices of international degree programmes in Finnish higher education. In A. Doiz, D. Lasagabaster, & J. M. Sierra (Eds.), *English-medium instruction at universities: Global challenges*. De Gruyter. (pp. 131–150). https://doi.org/10.21832/9781847698162-011

Sah, P. K., & Li, G. (2018). English medium instruction (EMI) as linguistic capital in Nepal: Promises and realities. *International Multilingual Research Journal*, *12*(2), 109–123. https://doi.org/10.1080/19313152.2017.1401448

Sainsbury, M., & Clarkson, R. (2008). *Attitudes to reading at ages nine and eleven: Full report*. NFER.

Salö, L., & Josephson, O. (2013). *Parallellspråkighet vid svenska universitet och högskolor* [Parallel language use at Swedish universities]. Stockholm University.

Salö, L., & Josephson, O. (2014). Landrapport Sverige. Parallellspråkighet vid svenska universitet och högskolor [Country Report Sweden. Parallel language at Swedish universities and colleges]. In F. Gregersen (Ed.), *Hvor parallelt. Om parallellspråkighet på Nordens universitet* [How parallel: On parallel language use at Nordic universities] (pp. 261–322). Nordic Council of Ministers.

Shrestha, P. N. (2009). ELT, ESP & EAP in Nepal: Whose interests are served? In M. Krzanowski (Ed.), *EAP and ESP in developing countries: State of play vs. actual needs and wants* (pp. 191–210.). Garnet Publishing.

Söderlundh, H. (2012). Global policies and local norms: Sociolinguistic awareness and language choice at an international university. *International Journal of the Sociology of Language*, *216*, 87–109. https://doi.org/10.1515/ijsl-2012-0041

Taylor, L. (2006). The changing landscape of English: Implications for language assessment. *ELT Journal*, *60*(1), 51–60. https://doi.org/10.1093/elt/cci081

Wächter, B., & Maiworm, F. (2008). *English-taught programmes in European higher education: The picture in 2007*. Lemmens Medien.

Wächter, B., & Maiworm, F. (2014). *English-taught programmes in European higher education. The state of play in 2014*. Lemmens Medien.

Weir, C. J., Hawkey, R., Green, A., & Devi, S. (2012). The cognitive processes underlying the academic construct as measured by IELTS. In L. Taylor & C. J. Weir (Eds.), IELTS collected papers 2: Research in reading and listening assessment: Vol. 34. Studies in language testing (pp. 212–269). Cambridge University Press.

Appendix A

Participant Demographic Information

Please provide the following information:
Age

Gender Male c Female c

Nationality

Languages used

Please list the languages you regularly use **for speaking** in order of frequency of use.
Please list the languages you regularly use **for writing** in order of frequency of use.
Please list the languages you regularly use **for reading** in order of frequency of use.
Name of the program you are studying on

Year of study
Level Bachelor □ Master □

Department and faculty

Student status Home student □ Visiting student □ Exchange student □
Other □ (please specify on the line below)

Any other relevant information

Would you like to take part in a follow-up interview this week?

You will be reimbursed with a voucher of SEK 100 if you take part.

Yes □ No □

Appendix B

Student Interview Protocol

Questions about you and your program of study

- What's your name?
- What language(s) do you use regularly and for what purposes?
- What program/qualification are you studying on?
- What made you decide to study on this program?
- Are you happy with your choice of study? Why/why not?
- To what extent is English used on the program? By teachers, students, yourself?
- How much of the course and assessment material is in English?

Your experiences of studying in an EMI environment

- What, in your view, are the main differences between studying courses through English and studying courses through your first language?
- What, in your view, are the advantages and disadvantages of studying in English compared to studying in your first language?
- If you could decide which language to study in, what would you choose, and why?

Your experiences of reading course material in English

- How much of the course material you read is in English?
- Does it make a difference to your understanding of a text which language it is in? How/why?
- On a scale of 1 – 10, how would you rate your overall understanding of the course material in English?

- What are your strategies for managing any difficulties? E.g. consulting dictionaries, noting difficult words in a vocabulary log, guessing from the context, asking someone, ignoring it, abandoning the text, or anything else?

Your experiences of taking the TOEFL reading test

- Having taken the TOEFL reading test, to what extent do you feel that it reflects the reading skills needed for a university course where English is the medium of instruction?
- To what extent do you feel that the TOEFL test is an accurate reflection of your English reading ability?
- Overall, how would you rate your understanding of the reading material in the TOEFL test?

Researching Academic Reading in Two Contrasting English as a Medium of Instruction Contexts at a University Level.

Suggested citation:

Owen, N., Shrestha, P. N., Hultgren, A. K. (2021). *Researching academic reading in two contrasting English as a medium of instruction contexts at a university level* (TOEFL Research Report No. RR-91). ETS. https://doi.org/10.1002/ets2.12317

Action Editor: John Norris

Reviewers: This report was reviewed by the Research Subcommittee of the TOEFL Committee of Examiners.