ETS TOEFL

Quality Beyond Measure.

TOEFL iBT® Research Report
TOEFL iBT–30
ETS RR–17-41

Investigating the Predictive Validity of *TOEFL iBT*® Test Scores and Their Use in Informing Policy in a United Kingdom University Setting

Claudia Harsch

Ema Ushioda

Christophe Ladroue

December 2017

The *TOEFL*® test was developed in 1963 by the National Council on the Testing of English as a Foreign Language. The Council was formed through the cooperative effort of more than 30 public and private organizations concerned with testing the English proficiency of nonnative speakers of the language applying for admission to institutions in the United States. In 1965, Educational Testing Service (ETS) and the College Board assumed joint responsibility for the program. In 1973, a cooperative arrangement for the operation of the program was entered into by ETS, the College Board, and the *Graduate Record Examinations*® (*GRE*®) Board. The membership of the College Board is composed of schools, colleges, school systems, and educational associations; GRE Board members are associated with graduate education. The test is now wholly owned and operated by ETS.

ETS administers the TOEFL program under the general direction of a policy board that was established by, and is affiliated with, the sponsoring organizations. Members of the TOEFL Board (previously the Policy Council) represent the College Board, the GRE Board, and such institutions and agencies as graduate schools of business, 2-year colleges, and nonprofit educational exchange agencies.

❖ ❖ ❖

Since its inception in 1963, the TOEFL has evolved from a paper-based test to a computer-based test and, in 2005, to an Internet-based test, the *TOEFL iBT*® test. One constant throughout this evolution has been a continuing program of research related to the TOEFL test. From 1977 to 2005, nearly 100 research reports on the early versions of TOEFL were published. In 1997, a monograph series that laid the groundwork for the development of TOEFL iBT was launched. With the release of TOEFL iBT, a TOEFL iBT report series has been introduced.

Currently this research is carried out in consultation with the TOEFL Committee of Examiners (COE). Its members include representatives of the TOEFL Board and distinguished English as a second language specialists from academia. The committee advises the TOEFL program about research needs and, through the research subcommittee, solicits, reviews, and approves proposals for funding and reports for publication. Members of the TOEFL COE serve 4-year terms at the invitation of the Board; the chair of the committee serves on the Board.

Current (2017-2018) members of the TOEFL COE are:

Lia Plakans – Chair	**The University of Iowa**
Yuko Goto Butler	University of Pennsylvania
Sara Cushing	Georgia State University
Ayşegül Daloğlu	Middle East Technical University (METU)
Luke Harding	Lancaster University
Claudia Harsch	University of Bremen
Lianzhen He	Zhejiang University
Volker Hegelheimer	Iowa State University
Lorena Llosa	New York University
Carmen Muñoz	The University of Barcelona
Marianne Nikolov	University of Pécs
Randy Thrasher	International Christian University
Paula Winke	Michigan State University

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 nonprofit Charter and Bylaws, ETS has and continues to learn from and 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 iBT Research Report Series and ETS Research Report Series ISSN 2330-8516

RESEARCH REPORT

Investigating the Predictive Validity of *TOEFL iBT*® Test Scores and Their Use in Informing Policy in a United Kingdom University Setting

Claudia Harsch,[1] Ema Ushioda,[2] & Christophe Ladroue[3]

1 University of Bremen, Bremen, Germany
2 Centre for Applied Linguistics, University of Warwick, Coventry, UK
3 Department of Computer Science, University of Warwick, Coventry, UK

The project examined the predictive validity of the *TOEFL iBT*® test with a focus on the relationship between TOEFL iBT scores and students' subsequent academic success in postgraduate studies in one leading university in the United Kingdom, paying specific attention to the role of linguistic preparedness as perceived by students and tutors. We employed a mixed-methods approach to enrich traditionally quantitatively oriented studies with a qualitative perspective. For the sample of 504 students who entered the university for postgraduate studies in the years 2011–2013 on the basis of a TOEFL iBT score, we analyzed the relation between TOEFL iBT scores and final academic award by correlation and regression analyses, taking into consideration discipline, nationality, and additional language support. For the qualitative strand, students entering the university in 2013 on the basis of a TOEFL iBT score were invited to complete questionnaires and interviews, as were their English for academic purposes and academic tutors. A total 48 students and 58 tutors participated, with 25 students and 36 tutors being interviewed at 3 points over the course of the year. Our findings show that students entering the university on the basis of TOEFL iBT scores feel well prepared and generally regard the test as an effective means of preparation for their academic studies in a U.K. setting. They cope well with linguistic demands, and a vast majority graduate successfully. Our findings support the appropriateness of the university's entrance policy with regard to setting minimum test score requirements, thus underpinning the predictive validity of TOEFL iBT in a U.K. setting.

Keywords Predictive validity; linguistic preparedness; mixed-methods study; regression analyses; accompanying students over course of one year; students' and their tutors' perceptions; role of language for academic success

doi:10.1002/ets2.12167

The *TOEFL iBT*® test is a relatively recent development, having been introduced worldwide only in 2006 (Cho & Bridgeman, 2012). Because it is a new test, validation of its proposed score interpretation and test use is paramount. The validation process for TOEFL iBT began with the conceptualization and design of the test (Educational Testing Service [ETS], 2008). It makes use of Kane's (2006) framework of a validity argument to synthesize evidence to support the proposed interpretations and uses of TOEFL iBT. There is by now a substantial body of evidence published (see, e.g., Chapelle, Enright, & Jamieson, 2008, or validation research published in the ETS Research Reports Series[1]), with the majority of research having focused on the North American context. In Kane's validity argument framework, certain claims are proposed with regard to specific test uses, and each claim needs to be supported by research evidence. Two of the proposed test uses are of relevance for our study: the use of TOEFL iBT for university admissions and for placement decisions with regard to language support programs. TOEFL iBT scores are meant to predict whether test takers have the English-language ability needed to be successful in an academic program. The proposed research is designed to provide evidence relevant to the following claim:

> The test score reflects the ability of the test taker to use and understand English as it is spoken, written and heard in English-medium college and university settings. The score is useful for aiding in admissions and placement decisions, and for guiding English-language instruction. (Enright et al., 2007, p. 6)

This proposition needs backing up by evidence in terms of the "relationships between test scores and. .. academic placements" (ETS, 2008, p. 3); evidence is also needed to back up the claim that the test "discriminates between students who

Corresponding author: C. Harsch, E-mail: harsch@uni-bremen.de

do or do not require additional language training" (Enright et al., 2007, p. 18). An initial positive relationship between TOEFL iBT scores and academic placement (with regard to language support programs or direct entry into academic studies without linguistic support) was found in the field study reported by Wang, Eignor, and Enright (2008).

Since TOEFL iBT differs considerably from the previous *TOEFL*® test versions and research to date has focused on the U.S. context (e.g., Cho & Bridgeman, 2012), further evidence is needed beyond the field tests to empirically underpin the use of TOEFL iBT scores in contexts outside North America. In the study reported here, we focus on the use of TOEFL iBT as one of the SELT[2] for U.K. university admissions purposes and for placement decisions regarding English-language support programs in one leading British university.

The continued growth in the international student population in higher education (HE) and the integration of increasing numbers of international students studying alongside home students pose various pedagogical challenges and concerns (see, e.g., the collection of papers in Carroll & Ryan, 2005; also Trahar, 2007). At the selected HE institution, concerns have been raised in various departments and across university management about international students who seem to lack a sufficient level of English to cope with the demands of academic study at the university, despite meeting English-language admissions criteria. From a more positive perspective, there are concerns about the ability to provide an academic environment, support systems, and resources that will enable all students (home and international) to develop and thrive at their academic studies.

It is in the context of these concerns that we conduct our study, focusing in particular on the use of TOEFL iBT for university admissions purposes and for placement decisions regarding English-language support programs. While the gatekeeping role of English-language tests is gaining increasing importance at various points of entry to life, study, and work in the United Kingdom, the predictive validity of such tests (i.e., the degree to which they can predict performance levels in a target behavioral domain of language use) is by no means a straightforward issue. Not only is it an open question whether an overall score or detailed skills profiles are the better indicator, but it also remains a challenge to set the actual cutoff points for entrance to academic studies as well as for placement decisions for additional language support classes. All university admissions staff, as well as tutors for English for academic purposes (EAP) teaching on language support courses, have to deal with this complex issue, yet there is little understanding which cut scores would be most adequate for admission and placement decisions.

Our study addresses two main research topics. First, we investigate how TOEFL iBT test scores relate to success in English-medium academic content programs in a British university, that is, outside of North America. The university is ranked among the top 10 in the United Kingdom and is part of the Russell Group.[3] We are looking at taught postgraduate programs, that is, full-time master's programs with a 1-year duration, as is usual in the United Kingdom. Furthermore, we examine the uses of TOEFL iBT scores (combined with other academic information, such as presessional grades and academic grades) for setting academic admissions policies and determining placement in English-language support programs in the selected university.

The research has a sector-wide relevance and is anticipated to make an academic and intellectual contribution in the following ways:

- inform stakeholders, such as university admissions officers, of appropriate entrance levels with regard to TOEFL iBT overall and section scores and
- support EAP management in placement decisions for pre- and in-sessional language support with regard to TOEFL iBT overall and section scores.

The project aims at providing valuable information about TOEFL iBT scores and the predictive validity of TOEFL iBT scores on academic performance in a selected HE institution in the United Kingdom not only in terms of quantitative evidence but also from a qualitative perspective on how students and their tutors perceive linguistic preparedness for academic studies. By adding this qualitative perspective, the research aims at promoting greater awareness and understanding among EAP and academic tutors of the language needs of international students from different backgrounds.

Literature Review

In the U.K. context, the dominant SELT is undoubtedly the International English Language Testing System (IELTS), with the majority of predictive validity research focusing on this test. While insights can be drawn from research into other English-language tests used for university admissions and placement purposes, there is a need to investigate TOEFL iBT in

the U.K. context. In general, previous studies into the predictive validity of university entrance language test scores such as from TOEFL or IELTS have so far been inconclusive and in part contradictory. Several studies found that language entry test scores were not a good predictor of academic success (e.g., Cotton & Conrow, 1998; Dooey & Oliver, 2002). Other researchers found a moderate predictive effect (e.g., Ingram & Bayliss, 2007), whereas others found a positive relationship between test scores and academic performance (e.g., Feast, 2002; Hill, Storch, & Lynch, 1999; Huong, 2000; Kerstijens & Nery, 2000). In a study using logistic regression, Van Nelson, Nelson, and Malone (2004) found that TOEFL scores had a predictive effect on the academic performance expressed in grade point averages (GPA). More recently, Cho and Bridgeman (2012) conducted a quantitative large-scale study in the United States to investigate the relationship between TOEFL iBT scores and GPA. They found moderate yet meaningful correlations between the two indicators. Addressing the well-known issue that heterogeneous groups can veil correlation patterns existing for more homogeneous subgroups, Bridgeman, Cho, and DiPietro (2015) found higher correlations between TOEFL scores and GPA when grouping students by nationality and department as compared to the overall sample.

Given the inconsistent picture emerging from the literature, we conducted a quantitative small-scale pilot study into the predictive validity of IELTS as the dominant English-language test in the United Kingdom at the selected HE institution in 2010–2011. We briefly summarize the findings and implications of this study here; the internal report (Ushioda & Harsch, 2011) is available online.[4] We used readily available and self-reported quantitative data on IELTS scores and found that multiple linear regression analyses resulted in the best fitting regression model using the IELTS overall score and the IELTS writing score as independent variables and the final academic grade as dependent variable. The model predicted 33.6% of the variance, and it was a good fit for the data, $F = 21.97$, $df = 2$, $p < .001$, ß = 0.471. We also found that in general, IELTS scores rise with average academic grades. When examining the lower end of the academic grade scale, we found a weak indication that students coming in with IELTS 5.5 overall could not cope with their academic studies despite having attended a presessional course. However, as we only used quantitative data and focused on IELTS, we needed to further investigate this issue in relation to the second most popular SELT, that is, TOEFL iBT, and complementary qualitative data.

Qualitative approaches have been taken, for instance, by Bayliss and Ingram (2006) and by Paul (2007), who found test entry scores a valid prediction of students' linguistic behavior during their academic studies. The small numbers of participants in these studies, however, make it hard to generalize the findings (Bayliss & Ingram, 2006, looked at 28 students; Paul, 2007, investigated four cases). Therefore, further studies across different settings and/or larger samples are needed. Our study contributes insights from one U.K. institution, combining quantitative and qualitative approaches to yield rich data and to be able to triangulate qualitative findings on a quantitative background.

Reasons for the earlier mentioned inconclusive research findings can be found in the difficulties of designing predictive validity studies. As Cho and Bridgeman (2012) pointed out, the relationship between language proficiency and academic success is not a direct one, as there are many other factors influencing and determining academic success. There is no ideal criterion variable, with GPA being the most widely used due to the lack of a better option and its ease of access. Other possible criteria could be self-assessments of language abilities (but see Wall, Clapham, & Alderson, 1994, warning against the use of self-evaluation) or the use of teacher ratings, which may be difficult to obtain in a reliable way. There have also been studies on the relationship between tutors' perceptions of students' linguistic preparedness and test scores, for example, by Bayliss and Ingram (2006), who found a close relationship between tutors' perceptions and students' IELTS scores. Though preparedness for academic studies can be conceptualized as a complex constellation of interacting linguistic, cognitive, social, and psychological variables (e.g., Collentine & Freed, 2004; Segalowitz & Freed, 2004), our study deliberately focuses on linguistic factors contributing to preparedness in order to address one specific factor. Hirsh (2007) suggested using diagnostic tests to investigate linguistic preparedness, yet the scarcity of diagnostic tools is a serious drawback (e.g., Alderson, 2005). Hirsh's (2007) suggestion to use DIALANG proved impossible in past attempts at the selected university due to network and server issues. Hence we consider perceptions of students and their tutors on linguistic preparedness for academic studies as an important window to enrich purely quantitative score data, to enhance our understanding and interpretation of quantitative indices, such as correlations and regression coefficients. Therefore our study includes students' and their tutors' perceptions of students' linguistic preparedness.

With regard to correlations, Cho and Bridgeman (2012) mentioned the problem of *range restriction* (test takers who are not admitted to university are not included in predictive validity studies), which leads to underestimated correlations but can be addressed by statistical adjustments, which we incorporate into our analyses (e.g., Sackett & Yang, 2000; Wiberg & Sundstrom, 2009). Following Cho and Bridgeman's (2012) suggestion, we use complementary methods, such as

expectancy graphs, to further address the issue of understanding relationships that may not be evident from correlations alone (see the Methodology and Design section). Another issue is the test purpose of TOEFL and similar tests, that is, to attest a certain "linguistic threshold" in order to function in an academic setting; this purpose is different from an academic admissions test (attesting "academic readiness"), yet "the relationship between English proficiency and academic performance is of interest to test users—especially admissions officers—and of relevance in supporting the use of test scores for high-stakes admissions decisions," as Cho and Bridgeman (2012, p. 4) rightly argued.

TOEFL iBT is used not only for university admissions but also for decisions on placement of international students in English-language support courses. In this field, research findings are also inconclusive: Wang et al. (2008) found support for using TOEFL iBT for such placement decisions, whereas Fox (2009) found issues of misplacement when using TOEFL and IELTS for such placement in one Canadian university. Kokhan (2012) reported that placement accuracy also depended on the time lag between taking TOEFL iBT and being admitted to university; hence it is advisable to control the date when a test was taken when exploring the feasibility of placement decisions. In the U.K. context, test results are only accepted if the test is taken within the 2 years preceding university enrollment, so we do not expect major effects of time lag in our context, while we acknowledge that some differential growth can be expected over this interval.

Given the inconclusive research findings, there is a need to investigate the use of TOEFL iBT for admissions and placement decisions in specific local contexts outside the United States, where TOEFL may not be the most widely used test, to increase its acceptability by supporting admissions and placement policies with empirical evidence. It is in this context that we conduct our mixed-methods study, to enhance our understanding of how TOEFL iBT can be used for university admissions purposes from a quantitative as well as a qualitative perspective. The fact that a variety of factors influence academic success or failure (see, e.g., Bayliss & Ingram, 2006) leads us to concentrate on factors that can be trained or addressed to facilitate academic success, focusing on students' and tutors' perceptions of linguistic preparedness for academic studies. Though our study focuses only on certain accessible factors, we fully acknowledge the influence of other variables that are beyond the control of the present study.

Research Questions

Based on the research findings and our pilot study discussed earlier, the present study incorporates quantitative and qualitative data and aims to address the following overarching questions:

1. What minimum entrance scores can be recommended for selected academic disciplines for students to be equipped with the necessary language skills to function in postgraduate studies?
2. What recommendations can be made with a view to placing students (with certain TOEFL iBT overall and section scores and certain academic disciplines) in presessional and in-sessional linguistic support programs?

The project encompasses two strands with particular research aims. Strand 1 looks at readily available quantitative data, that is, TOEFL iBT scores, presessional grades, in-sessional attendance, and final academic grades, to examine the explanatory and predictive power of language proficiency entrance scores and language support on international students' academic success expressed in the final course work/academic grade. Within Strand 1, we address the following research questions (RQs) and subquestions:

RQ1. What is the relation between the language skills profiles reported by the TOEFL iBT section scores and students' subsequent academic performance as expressed in final academic grades?

a. Do different subgroups (e.g., nationality, subject discipline, additional language support) of students show differing profiles in their TOEFL iBT overall and section scores?
b. What are the relations between TOEFL iBT overall and section scores and academic grades for different subgroups?
c. What effect has additional language support on the final academic grade and on the relation between TOEFL iBT scores and the final academic grade?

RQ2. What is the predictive potential of the TOEFL iBT scores with regard to predicting students' final academic grades?

a. What is the predictive potential of the TOEFL iBT overall score and the TOEFL iBT section scores on students' final academic grades?
b. Do selected variables (students' nationality, academic disciplines, additional language support) have an effect on the predictive relation between TOEFL iBT scores and academic outcome?

Strand 2 uses a mixed-methods approach to enrich the quantitative data from Strand 1 with quantitative and qualitative self-report data (questionnaires and interviews) to investigate the predictive and explanatory power of TOEFL iBT scores[5] in students' and their tutors' perceptions of students' linguistic preparedness for academic studies. Moreover, we explore students' and their tutors' perceptions of students' exploitation of language support and of the effectiveness of the support offered before and during studies. Another perspective is the perceived role and influence language has on academic progress. Within Strand 2, we address the following specific RQs and subquestions:

RQ3. What role does TOEFL iBT play in students' and their tutors' perceptions of students' linguistic preparedness for academic studies?

a. Do students/tutors feel that students are prepared for/can cope with linguistic demands?
b. Do students/tutors think that TOEFL iBT prepared students well for/is a good predictor of preparedness and academic success?
c. Do students' perceptions change over the year?

RQ4. How do students exploit language support, and what are the links between students' language weaknesses (as perceived by students/tutors or as reported by TOEFL iBT) and seeking support?

a. What support do students need, seek, and exploit, as perceived by students and their tutors, and what are their reasons for (not) seeking support? Do students and tutors think the support is effective?
b. What is the relation between linguistic struggles/weaknesses (as perceived by students/tutors or as reported by the TOEFL iBT report) and seeking support?

RQ5. What role does language play in academic success?

a. What role does English-language proficiency play in academic success and assessment/feedback practices across selected departments/faculties, as perceived by students and their tutors?
b. What effect does students' English-language proficiency have on their academic success?

Methodology and Design

The interdisciplinary research employs a mixed-methods approach, involving applied linguists and statisticians. Ethical approval was sought following university regulations; the research project was fully approved by the Humanities and Social Sciences Research Ethics Committee of the university on July 31, 2013. In what follows, we describe our design, the sample of participants, the instruments used, the data collection procedures, the variables used, and analyses conducted separately for the two strands and the respective research questions. We also explain how the answers to the research questions relate to particular components of the validity argument associated with TOEFL iBT.

Strand 1

Data Set

For Strand 1, we made use of readily available quantitative data sets provided by Central Registry at the university where the research was conducted. Central Registry routinely collects the following data for all postgraduate students entering the university with a SELT ($N = \sim$1,500 per annum; $n = 180–200$ for TOEFL iBT test takers): scaled TOEFL iBT section scores, date when the test was taken, demographic background data (age, gender, first language), degree chosen, department, and final academic grade reported for the obtained degree.

With regard to the final academic grade, we have to concede that individual departments across the university have different assessment practices to achieve the final grade. Yet each faculty employs an agreed set of marking criteria, using the same university-wide 100-point marking scale for assessing course work, aiming at reliable assessment and marking procedures. The standards and procedures employed at the selected university are accepted practice across the United Kingdom.[6] Across the university, the final academic grades are based on marked course work and dissertation or project completion as well as on attendance. Depending on the faculty and the academic discipline, different assessment approaches, assignment formats, and sets of marking criteria are in use. Marking criteria focus on aspects such as critical analysis, application of knowledge, or understanding; there is, generally, one criterion focusing on language aspects, such as communication (for oral assignments) or presentation (for written assignments). The extent to which language determines the academic course work grade differs across faculties and disciplines; hence we used Strand 2 to examine different practices in different academic disciplines and the perceived influence of language on academic success.

To obtain the final academic grade, the 100-point marking scale is divided into four grades in the following way: below 50 = fail; 50 – 64 = pass; 65 – 69 = merit; 70 and above = distinction. There is also the possibility of obtaining a lower than intended degree (e.g., a postgraduate diploma rather than a master's degree) rather than failing the course under certain circumstances, usually with marks in the range of 40 – 49. Because U.K. universities regularly report the final grade only with reference to these five degree classifications, we used the following five ordinal grades for our analyses: *fail*, *lower degree*, *pass*, *merit*, and *distinction*.

In addition to the data from University Admission, we also used data on language support programs run by the university's Centre for Applied Linguistics, that is, presessional grades and in-sessional attendance. Students who enter with a slightly lower than required TOEFL iBT (the thresholds vary across the disciplines) are placed in the presessional program, lasting either 5 or 10 weeks, depending on the TOEFL iBT scores. Attendance of the presessional classes is compulsory and monitored, as is completion of course work tasks, which are assessed continuously for formative purposes; no formal grades are collected for the ongoing course work. Data on presessional exit grades are obtained via standardized in-house tests at the end of the course, targeting the four linguistic skills (reading, listening, writing, speaking), which are marked with reference to an agreed set of criteria that all tutors use; test results are reported on a 4-point grading scale (fail/pass/merit/distinction). With regard to in-sessional courses, students can opt at the beginning of a term for a one-term in-sessional program (consisting of 10 classes); attendance of classes is monitored but optional, and no assessment takes place. Hence attendance data were aggregated over all classes and terms to obtain the total hours of class attendance over the year.

To sum up, the data set used for Strand 1 contained scaled TOEFL iBT overall and section scores (interval-scaled variables), demographic data (department, degree chosen, age, gender, first language; nominal variables), presessional and in-sessional attendance (if attended; nominal variable), and final academic grade (ordinal variable).[7]

Sample

Strand 1 of this study includes all students having entered the university on the basis of TOEFL iBT scores in the years 2011, 2012, and 2013 for postgraduate studies (the majority attending a 1-year taught master's course), excluding students who withdrew from their courses. This TOEFL iBT sample comprised 483 students altogether, of whom 74 attended additional language support classes (for a detailed breakdown, see the Data Summary and Descriptive Statistics subsection).

Methods of Analysis

To answer RQ1, we used simple plots and cross-plots as exploratory tools (e.g., French, 2011; Tukey, 1977) as well as correlation analyses (e.g., Gliner, Morgan, & Leech, 2009; Krzanowski & Marriot, 1994, 1995). To address the issue of range restriction (our sample only entails students with TOEFL iBT scores above a certain threshold), we used statistical correction formulas (e.g., Sackett & Yang, 2000; Wiberg & Sundstrom, 2009; we used Thorndike Case 2, as will be explained later). In a first step, we examined whether different subgroups of students showed differing profiles in their TOEFL iBT section scores. Controlling score profiles seemed necessary, because Bridgeman et al. (2015) identified a subgroup of students sharing a conspicuous profile in their TOEFL iBT section scores, which was distorting the reported correlation. We investigate whether our students showed similar distorting patterns before we conducted correlational analyses.

In a second step, we examined the relations for different subgroups of our sample. Bridgeman et al. (2015) noticed that using correlation on the whole sample is a rather blunt approach that could lead to misleading results. Subgroups can reveal differing relations that may be lost when the whole sample is considered. This is an example of Simpson's paradox (Wagner, 1982), in which an apparent negative relationship between two variables turns into a positive relationship once subgroups are taken into account. We considered a number of different groupings that could influence the relationship between students' TOEFL iBT score profiles and their final academic grades, such as faculties, disciplines, and nationalities.

Finally, we examined the effect additional language support could have on the final academic grade and on the relation between TOEFL iBT scores and the final academic grade. Some students, particularly those with lower TOEFL iBT scores, have to attend additional language support classes in advance of the academic studies, so-called presessional classes. Other students attend these classes voluntarily, and yet other students attend in-sessional classes on a voluntary basis during their studies. We investigated what effect these classes have on the final academic grade, while taking the students' TOEFL iBT scores into account.

We then addressed RQ2 to examine, on one hand, the predictive relation between the TOEFL iBT overall score and the TOEFL iBT section scores and students' final academic grades and, on the other hand, to explore whether selected variables (students' nationalities, academic disciplines, additional language support) have an effect on the predictive relation between TOEFL iBT scores and academic outcome.

For the TOEFL iBT sample, we explored the explanatory power of the following indicators (independent variables) on academic success (dependent variable as expressed by final academic grades): TOEFL iBT scores, departments, students' first languages as indicated by their nationalities, and additional language support. The main predictive variable (TOEFL iBT scores) was interval scaled, whereas the outcome variable was ordinal (five degree classifications of fail, lower degree, pass, merit, and distinction). Given the ordinal outcome variable and the earlier mentioned restriction of range in the predictive variable, our data set was of only moderate variance. We addressed this challenge by first employing expectancy graphs as an efficient way of summarizing the data, as suggested by Cho and Bridgeman (2012), to explore the probability of obtaining a certain academic grade given a certain range of TOEFL iBT scores. In a second step, we examined the power of TOEFL iBT scores as predictors of final academic grades by using an ordered logistic regression (OLR; Agresti, 2002), in which we fit a range of models to model the predictive relation between TOEFL iBT scores (and additional selected other variables, such as nationality and discipline) and the final academic grade. In the same way that a linear regression fits the best line between the variables to predict the outcome, an OLR uses a procedure for estimating the probabilities of belonging to each of the outcome's categories (the final academic grade in this context). This was achieved by mapping the result of a linear fit of the transformed predictive variables to the outcome's categories. As reported by Bridgeman et al. (2015), correlations can be difficult to interpret. An advantage of OLR is that the probability of an event can be derived from the regression coefficients, which makes the results easier to interpret. A regression is also useful for addressing the aforementioned Simpson's paradox by automatically including an interaction between different predictors.

Because we did not rely on correlational analyses alone, the techniques we employed (i.e., expectancy graphs and OLR analysis) made it possible to predict the final academic grade based on TOEFL iBT scores. The findings contribute to empirically underpin the classical predictive validity argument (i.e., predicting the probability of achieving a certain final academic grade based on a certain TOEFL iBT score) also in terms of explained variance in final grades by TOEFL iBT scores and the probabilities of academic success predicted by TOEFL iBT scores.

All statistical analyses were run with the program R (R Core Team, 2016).

We are aware of the drawbacks of a purely quantitative approach as outlined here (such as no access to the full range of test takers or factors influencing academic success that could not be controlled) but nevertheless regard the preceding detailed quantitative analyses as a reliable and necessary backdrop on which to interpret and discuss our qualitative findings from Strand 2. We regard this mixed-methods approach as one possibility to triangulate our findings.

Strand 2

Design

For Strand 2, we focused on the cohort of students entering the university on the basis of TOEFL iBT scores in 2013 who participated voluntarily in a survey via online questionnaires and follow-up interviews. We collected quantitative and qualitative data during the academic year 2013–2014 (for different data collection points, see the Sample and Data

Table 1 Design of Strand 2

Beginning Term 1	Beginning Terms 2 and 3	End Term 3
Students		
1. Online questionnaire about student perception of being prepared for academic studies	2. First and second interviews 3. Nominate tutors to be interviewed	4. Online questionnaire 5. Third interview 6. Provide contact to dissertation supervisor
Tutors		
1. Online questionnaire on student preparedness, perception of TOEFL, and student support	2. First and second interviews with nominated tutors	3. Interview with dissertation supervisor

Collection subsection). We employed a longitudinal design to accompany students throughout the academic year, with questionnaires for students at the beginning and end of the academic year and interviews at three points throughout the academic year. The student perspective was complemented by the perspective of EAP and academic tutors, who were invited to participate in a survey. Moreover, students participating in our interviews were asked to nominate tutors and their dissertation supervisors to be interviewed, also at three points throughout the academic year. This angle served to compare perceptions of preparedness reported by students and tutors and to triangulate findings derived from the students' perspectives by mapping themes found in both perspectives. Table 1 gives an overview of the design.

Questionnaires and follow-up interviews are regarded as adequate instruments to collect survey data (Gass & Mackey, 2007): Questionnaires allow a practical and flexible implementation; they offer the possibility to reach a larger sample and to administer the same set of questions (Dörnyei, 2010). Drawbacks, such as missing important aspects, to questionnaires can be addressed by the follow-up interviews, which are conducted in a semistructured way (Kvale, 2007; Merriam 2009). Surveys allow for systematic data collection and analysis (Cohen, Manion, & Morrison, 2011) and for combining quantitative and qualitative aspects. This combination offers opportunities for elaboration, explanation, and confirmation of data (Jang, McDougall, Pollon, Herbert, & Russell, 2008), thus supporting the understanding and interpretation of outcomes (Bryman, 2008).

A survey-based approach allowed us to investigate in more depth the earlier indicated issue that marking procedures vary across the university, its faculties, and its academic disciplines. We examined general assessment practices, criteria used, and the influence of language on the academic course work grades across different disciplines via questionnaires. This was complemented by interviews, in which we examined the tutors' and students' perceptions of the importance of particular language skills for particular disciplines. This perspective could be used to enhance and inform decisions on discipline-specific cut scores currently in use at the university.

Instruments

Questionnaires

With regard to the constructs targeted in the student and tutor questionnaires, we examined preparedness in terms of being able to cope with linguistic demands during academic studies after having taken TOEFL iBT. We looked at all four skills and covered the demands arising from the following common activities in an academic setting (based on but not restricted to the linguistic behavior categories in Bayliss & Ingram, 2006): listening to lectures, seminars, and tutorials; taking notes while listening; reading and processing background literature; writing assignments (the most common type of module assessment); preparing and giving presentations; and group work. Given our specific research focus on examining the relevance of TOEFL iBT to university settings outside the United States, we felt it was particularly appropriate to draw on previous analyses of academic tasks and associated language skills in non-U.S. settings rather than refer to the TOEFL iBT research literature, such as Rosenfeld, Leung, and Oltman (2001), where the analysis has focused on North American university contexts. We also asked students whether they has prepared for taking TOEFL iBT, and we collected information about the means of preparation[8] as well as the perceived effectiveness of it (partly based on O'Loughlin, 2008). Furthermore, the questionnaire encompassed students' and tutors' views on how well TOEFL iBT covers the linguistic demands encountered in an academic setting, and hence how well students and tutors feel the test can measure students'

readiness for academic studies.[9] With regard to exploiting language support at the selected university, we asked students whether and which support they have exploited/are exploiting and their perceptions of effectiveness (informed in part by O'Loughlin, 2008); we asked tutors what support is in place in their departments and across the university to support international students.[10] Our questionnaires also asked for background data, including requirements to attend pre- or in-sessional courses, and exploitation of language support at the university (in part based on O'Loughlin, 2008; Woodrow, 2006). We used a combination of closed items (multiple choice and Likert-type rating scales) and open questions. The questionnaires are accessible online.[11]

The questionnaires were piloted with small convenience samples: We received extensive feedback from five PhD students at the university on the mainly qualitative student questionnaires, which was used to revise the final student versions. With regard to the tutor questionnaire, we piloted the mainly qualitative tutor questionnaire within our network of academic lecturers and EAP tutors outside the university. Following common practice in the U.K. HE context, we used the term *tutor* to refer collectively to all staff involved in teaching students (whether academic subjects or English support classes), except where we wished to distinguish between "academic lecturers" and "EAP tutors." The extensive feedback we received from five tutors was used to revise the questionnaire.

Interviews

For the student and tutor interviews, we took up the main themes from the questionnaires, building on the questionnaire items and expanding them. This allowed us to link questionnaire and interview data. The targeted themes and interview questions were organized in interview guides (see Appendices A – G), giving guidance for the different interviewers involved in the project while at the same time allowing for flexibility to follow up on ideas brought up by our participants.

With regard to the first round of interviews, we covered four themes in the student interviews: how well students feel prepared linguistically and how they are getting on with the English-language requirements; students' perceptions and experiences of assessment procedures and the role of English at their departments; students' perceptions of the usefulness of TOEFL iBT test reports; and students' current exploitation of support with their English language. The first tutor interviews were structured parallel to the student interviews; we covered the same four aspects: the tutor's perceptions of how a particular student is prepared for and getting on with the English-language requirements, of TOEFL iBT test reports and their usefulness, of the role English plays in departmental assessment procedures and academic progress, and of language support on offer in general and for the particular student.

In the second round of interviews, we built on the themes of the first round, expanding our focus on the following four themes: how well the students are coping with the linguistic demands of their academic studies and whether they receive any feedback on their language; the students' perceptions of the relation between their TOEFL iBT scores and their academic assignment marks; the students' perceptions of the relation between any language support programs they attended/are attending and their academic assignment marks; and students' exploitation of support with their English language. With regard to the tutor interviews, we used the guide for the first interviews with all new tutors to capture their perceptions of the usefulness of TOEFL iBT score reports, of the role English plays in assessment and academic progress, and of language support provision, besides asking them for their perception of how the particular student is coping with the English-language demands of his or her academic studies. Tutors who participated in Round 1 were only interviewed with regard to their perceptions of how the student is currently getting on and coping with the English-language requirements.

In the third and final round of interviews for the students, we used an interview guide that builds on and expands the questions used in the second questionnaire. We invitde the students to reflect on how well they coped with the English-language requirements, how well they think TOEFL iBT prepared them linguistically for their studies, and how well the TOEFL iBT scores reflect their academic progress and success. We also asked students whether they think that their language proficiency and the fact that English is not their first language affected their academic progress. Furthermore, we explored with the students whether their English improved and what role any language support exploited may have played in their academic progress.

At the third interview point, we invited the students' dissertation supervisors, who closely worked with the students on a research project and on a written dissertation over a 3-month period. The vast majority of supervisors were new to our project. Hence we interviewed them on the following themes, focusing on a supervisor's perceptions of the student's English proficiency and how well the student was prepared for and coped with the academic linguistic requirements;

Table 2 Strand 2 Sample

Participants	Questionnaires	Interviews			
		All	1	2	3
48 students	Q1: 31; Q2: 19 (overlap: 8 students)	25	25	19	21
58 tutors (9 EAP/49 academic)	32	36	27	9	10

marking procedures and the role English plays in marking dissertations; how well TOEFL iBT test reports reflect the student's academic achievements; and language support offered during the dissertation writing stage for the particular student and in general.

Sample and Data Collection

For Strand 2, we invited all 223 students entering the university in 2013 on the basis of a TOEFL iBT score via Central Registry to participate voluntarily in our study. Information sheets and consent forms were provided as required by the university's ethics regulations. To secure a high participation rate, we offered a prize drawing for the questionnaire participation (as is common practice in the United Kingdom) and a small compensation for participation in the follow-up interviews. Students were first invited to fill in the questionnaire, where they could indicate their interest in further participation. Those indicating their interest were then contacted and invited for the interviews. Students participating in the interviews were asked at each point whether they would be willing to nominate one or more tutors to be interviewed, including their dissertation supervisor at Interview 3. The nominated tutors were then invited to the interviews on a voluntary basis. Meanwhile, all EAP tutors at the university and all academic lecturers at the departments with the biggest intake of postgraduate students were invited to participate in the staff survey on a voluntary basis. At the end of the academic year, Central Registry again invited all students to participate in the second questionnaire. Table 2 depicts the final sample of participating students and tutors for Strand 2.

With regard to the different instruments and stages of our study, a complex picture of participation emerges, because not all participants took part in all stages of the study. For instance, we only had one student who filled in both questionnaires and participated in all the interviews. Seventeen students who participated in the interviews also filled in Q1, 7 of whom also filled in Q2. We had eight students who filled in both questionnaires, seven of whom also participated in the interviews. Overall, eight interviewed students also filled in Q2. With regard to the tutors, we had 10 tutors who filled in the questionnaire and participated in the interviews.

With regard to interview participants, the column "all" represents the actual number of persons who participated in the interviews; some participated only in one interview, others in two, and some in three interviews. In total, 25 students participated in the first interview, 19 of whom returned for the second interview, and 21 of whom took part in the third. With regard to the tutors, all in all, 36 tutors participated, some of whom in two or more interviews; hence the number of participants in the "all" column does not present the sum of the three interview columns. Because all participants can be identified via project IDs, we could link questionnaire and interview data to individual participants. Moreover, via the student ID collected during the interviews, we could link interviewed students from Strand 2 to the data set from Strand 1 to draw on the quantitative data from Strand 1.

In total, we conducted 111 interviews, 65 with students and 46 with tutors. The student interviews lasted between 11 and 52 minutes, amounting to a total of 24 hours of recordings. The interviews with the tutors lasted between 5 and 65 minutes, with a total of 19 hours of recordings. The recordings were transcribed by a professional service and the transcripts imported into NVivo software for qualitative analysis.

Methods of Analysis

In Strand 2, wed use a combination of quantitative and qualitative analysis methods. The questionnaire data were analyzed by descriptive statistics to gain quantitative insight into students' and tutors' perceptions of preparedness and language support exploitation. The qualitative data from the questionnaires and interviews were analyzed by a combination of deductive and inductive approaches (Cohen et al., 2011; Dörnyei, 2007), such as directed and thematic analysis (Braun

& Clarke, 2006), employing NVivo software. They complement and inform the quantitative insights on preparedness and language support exploitation.

Our three main RQs and related subcategories, as well as all variables (items) from the questionnaires and interviews, are shown in Appendix H for student instruments and in Appendix I for tutor instruments. Reading Appendices H and I across the rows gives an overview of how the variables feed into our research themes and questions. To qualitatively analyze the interview data and open comments from the questionnaires, we developed a coding scheme based on Appendices H and I. The coding scheme was revised in eight iterative cycles, with the interview data informing the revisions in an inductive way. Appendix J shows our final coding frame, which reflects the research themes and research questions. The data were coded by four coders, the principal investigator (PI), one co-investigator (Co-I), and two research assistants. (The PI, Co-I, and one assistant have PhD degrees in the social sciences, one assistant was working on her PhD degree, and all four are well versed in qualitative analysis.) As part of the coder training, the PI and Co-I initially coded several interviews and discussed the results until agreement was reached. The research assistants then coded the same interviews, compared their results with the precodes provided by the PI and Co-I, and discussed results with the PI and Co-I; as a result, the coding scheme was amended where necessary, the interviews recoded, and the results discussed again. This was reiterated until agreement was reached. In the ensuing coding cycles, each interview was coded by at least two coders, and all discrepancies were discussed until agreement was reached.

We answered the three RQs and their subquestions by quantitatively and qualitatively analyzing the main themes of preparedness, exploitation of language support, and the role language plays. The analyses were initially conducted separately for each research instrument, and within each instrument separately for students and their tutors, before comparing and contrasting the two perspectives for triangulation purposes. To answer our three RQs in turn, we then drew on the results from analyzing the questionnaire and interview data, again using the variables and themes outlined in Appendices C and D.

Operationalizing the three RQs in our questionnaire items and interview questions, using the three main RQs as guiding themes in the coding scheme (Appendix J), and linking all questionnaire and interview variables to the three main RQ/themes (Appendices H and I) creates a coherent and transparent system of analysis, which in turn allows us to draw on relevant variables to answer our RQs.

Overarching Aims

While in a first step, each strand will be analyzed to answer the particular research questions outlined earlier, the overall aim for the final report is to bring the strands together to investigate the feasibility of existing cut scores for admission purposes for selected academic tracks and disciplines as part of TOEFL iBT predictive validity and usage, taking into account perceived linguistic preparedness for academic studies and the exploitation of academic language support systems available.

The overarching questions our research aims to address are as follows: (a) What minimum entrance scores can be recommended for students to be equipped with the necessary language skills to function in postgraduate studies? and (b) What recommendations can be made with a view to placing students (with a certain profile of TOEFL iBT scores) on presessional and in-sessional linguistic support programs? To answer these two questions, we used all data and findings from Strand 1 (cohorts 2011, 2012, 2013, $n = 504$) and Strand 2 (cohort 2013, $n = 48$ students and 58 tutors). We interpreted findings from Strand 2 in light of the results we found for the cohorts in Strand 1, as a backdrop to confirm any emerging trends from our analyses in Strand 2. A mixed-methods approach was employed here, drawing on Morse's (1991), Creswell's (2009), Creswell and Plano Clark's (2011), and Teddie and Tashakkori's (2009) mixed-methods designs: We adopted a sequential design, using Strand 1 (previous years) to analyze the Strand 2 cohort in light of the Strand 1 findings. We paid careful attention to triangulating qualitative and quantitative findings to see where the perspectives that they bring are mutually supportive.

Findings From Strand 1

This section reports the results from the quantitative Strand 1 of the project. Its aim is to investigate the relation between the TOEFL iBT test scores and academic outcomes as well as the potential of the TOEFL iBT score to predict a student's academic outcome and to identify other factors that contribute to this relation.

Table 3 TOEFL iBT Sample by Year and Faculty

Year	Arts	Medicine	Sciences	Social sciences
2011	10	3	53	90
2012	15	2	54	117
2013	11	0	48	80
Total	*36*	*5*	*155*	*287*

Table 4 Students Attending Language Support Programs

Year	Presessional	In-sessional
2011	18	10
2012	14	8
2013	14	12

Table 5 Student Numbers by Department (Top 10)

Department	*N*
Business School	153
Manufacturing Group	107
Economics	55
Politics and International Studies	32
Theater, Performance, and Cultural Policy Studies	30
Centre for Applied Linguistics	17
Mathematics Institute	15
Law School	13
School of Life Sciences	12
Sociology	8
Other	41

Data Summary and Descriptive Statistics

The total number of students having entered the university on the basis of a TOEFL iBT score for a 1-year postgraduate course between 2011 and 2013 is 504. For the purposes of our study, we removed the students who withdrew ($n = 21$) so that our sample for the analyses encompassed 483 students (Table 3). In our further analyses on the faculties level, we excluded the Medicine Faculty because the group ($n = 5$) was too small to produce meaningful results.

Within our TOEFL iBT sample, Table 4 shows the number of students who attended the presessional and in-sessional programs.

Our sample has an average age of 24.67 years with a standard deviation of 3.30 and a median age of 24 years. We have a fairly balanced gender distribution of 230 female and 253 male participants.

We now report the student numbers by discipline and nationality (Tables 5 and 6).

We also looked into subsampling by department and nationality, because Bridgeman et al. (2015) found interesting correlation patterns when breaking their sample down by discipline and nationality. However, given our sample size, such a breakdown yields reasonable subsample sizes for two departments only, as can be seen in Table 7, where student numbers for the three departments and the three largest nationality subgroups are displayed.

We now report descriptive statistics for the TOEFL iBT score distribution. The TOEFL iBT test scores are rather high, as expected given the truncated sample mentioned above, as Table 8 shows.

Students in our sample achieved mean scores above 25 out of 30 possible for the subskills and, on average, a total of above 104 out of a possible 120 scores overall, the variance being rather small. As was to be expected from the truncated sample, the data are not normally distributed, as indicated by the Shapiro–Wilk test (all p values $< .001$). To account for this, we used the Thorndike Case 2 correction formula (explained later in the Correlational Analyses subsection), which corrected the correlations for selection bias and the resulting ceiling effects.

Table 6 Student Numbers by Nationality (Top 10)

Nationality	*N*
Chinese	72
Indian	67
German	57
French	23
Turkish	23
Italian	15
Taiwanese	15
Thai	14
Mexican	12
Greek	11
Other	174

Table 7 Student Numbers by Top Three Departments and Nationality

	N
Business School	
Chinese	26
German	21
Indian	32
Manufacturing Group#	
Chinese	9
Indian	22
Turkish	15
Economics	
Chinese	7
German	11
Indian	5

Table 8 Scaled TOEFL iBT Scores, Descriptive Statistics

Scaled TOEFL iBT	Min.	Max.	Median	Mean	*SD*	Shapiro Wilk *p* value
Reading	10	30	27	26.77	2.73	<.001
Writing	17	30	27	25.99	2.58	<.001
Speaking	18	30	26	25.26	2.68	<.001
Listening	16	30	27	26.71	2.77	<.001
Overall	75	120	106	104.70	7.70	<.001

We now report descriptive statistics with regard to relevant subgroupings by selected faculties (Table 9), departments (Table 10), and nationalities (Table 11). We report only for subgroups with a substantial number of students.

Science students show a tendency for slightly lower scores, with a slightly higher variability, whereas social science students tend to have the highest scores in all areas but speaking, where arts students show the highest score values. There is a trend for science and social science students to achieve lower scores in speaking. Again, the Shapiro – Wilk tests confirmed that data are not normally distributed, which we accounted for by using the Thorndike correction formula.

Table 10 gives an overview of mean and standard deviation for the three largest departments: Business School, Manufacturing, and Economics. Business School students seem to achieve the highest scores with the lowest variance, while the Manufacturing group achieves relatively lower scores with the highest variance. Again, there is a trend for students across all three groups to achieve lower scores in speaking. For these groups, the Shapiro – Wilk tests also confirmed that data are not normally distributed (Table 11).

Indian students show a tendency for higher TOEFL iBT scores and smaller variance, whereas Chinese students show the lowest scores with a trend for the largest variance. While Chinese and German students in our sample show relatively

Table 9 Descriptives for Faculties

	Arts		Science[b]		Social sciences[c]	
TOEFL iBT	Mean	*SD*	Mean	*SD*	Mean	*SD*
Reading	26.56	2.46	25.97	3.26	27.21	2.32
Writing	26.67	2.24	25.10	2.85	26.43	2.33
Speaking	26.25	2.53	24.25	2.57	25.69	2.61
Listening	26.75	2.57	25.75	3.38	27.21	2.25
Overall	106.22	6.92	101.08	9.02	106.56	6.20

[a]$n = 36$. [b]$n = 155$. [c]$n = 287$.

Table 10 Descriptives for Departments

	Business School		Manufacturing Group[b]		Economics[c]	
TOEFL iBT	Mean	*SD*	Mean	*SD*	Mean	*SD*
Reading	27.67	1.95	25.51	3.43	26.65	2.63
Writing	26.97	2.01	25.02	2.92	25.80	2.63
Speaking	26.19	2.46	24.26	2.78	25.13	2.94
Listening	27.75	1.99	25.28	3.43	26.96	2.26
Overall	108.60	4.64	100.10	9.38	104.60	7.27

[a]$n = 153$. [b]$n = 107$. [c]$n = 52$.

Table 11 Descriptives for Nationalities

	Chinese[a]		Indian[b]		German[c]	
TOEFL iBT	Mean	*SD*	Mean	*SD*	Mean	*SD*
Reading	27.18	2.51	26.97	2.38	26.69	2.13
Writing	25.64	2.81	26.32	2.27	26.94	2.10
Speaking	23.35	2.10	27.00	2.02	26.31	2.48
Listening	26.15	2.74	27.97	1.96	26.88	2.56
Overall	102.32	7.34	108.35	4.86	106.82	6.37

[a]$n = 72$. [b]$n = 67$. [c]$n = 57$.

lower scores in speaking, this trend cannot be observed for the Indian group. As for these data, they are not normally distributed.

We now report correlations among the TOEFL iBT section and total scores using Pearson's product–moment coefficient for interval-scaled data (Table 12). All correlations are significant, with coefficients for the section scores ranging from .19 to .45. For the section scores, there seems to be a trend that the highest coefficients occur between the section scores for productive and receptive skills, respectively. With regard to coefficients for correlations between section scores and overall scores, they are all significant and range between .67 and .75. Given our sample and the earlier mentioned range restriction of the TOEFL iBT scores, there is enough variability in the score data to allow them to correlate with one another. Nevertheless, to account for this range restriction, we adjusted the correlations between TOEFL iBT scores and final academic grade using the Thorndike correction formula.

In what follows, we report the distribution of the final academic grades. For our total sample, Figure 1 shows a high pass rate (211), along with high numbers of merit (127) and distinction (124). We only had four fails in the sample and 17 lower degrees. Figure 2 displays the academic outcome broken down by faculty (Medicine omitted due to sample size).

Relations Between *TOEFL iBT*® Scores and Final Academic Grades

In this section, we report findings with regard to our first research question, that is, the relation between TOEFL iBT scores and final academic grades. However, before we could conduct correlational analyses, we had to check our data set for conspicuous profiles of TOEFL iBT section scores, because Bridgeman et al. (2015) reported a potential distortion of correlational results by imbalanced test score profiles.

Table 12 Correlations Among TOEFL iBT Scores

	Reading	Writing	Speaking	Listening	Overall
Reading	1	.39**	.19**	.44**	.71**
Writing	.39**	1	.45**	.34**	.75**
Speaking	.19**	.45**	1	.30**	.67**
Listening	.44**	.34**	.30**	1	.74**
Overall	.71**	.75**	.67**	.74**	1

*Correlation significant at .05, two-tailed. **Correlation significant at .01, two-tailed.

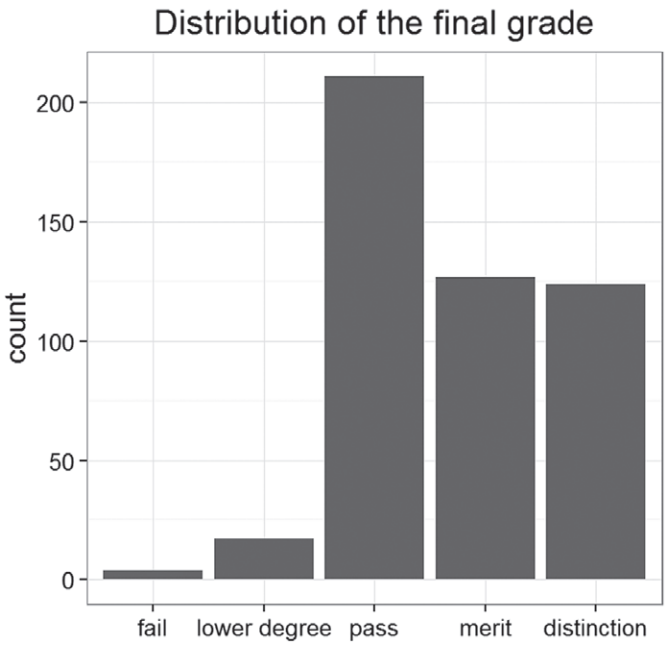

Figure 1 Distribution of final academic grades.

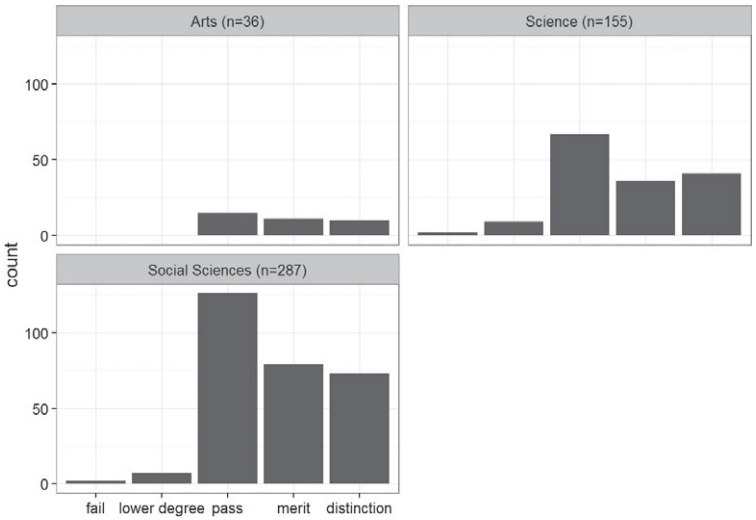

Figure 2 Academic outcome by faculty.

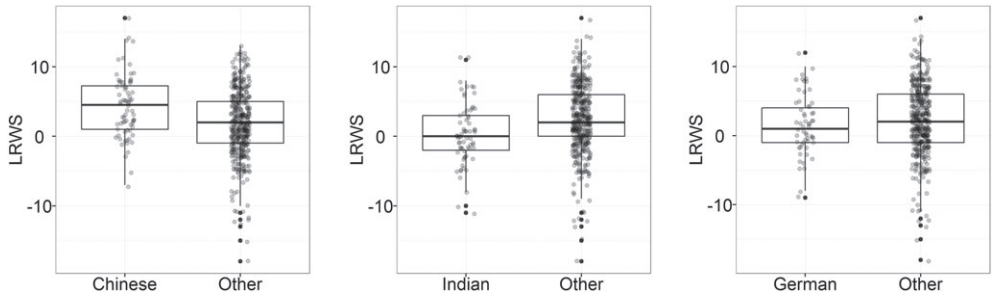

Figure 3 Box plots for (left) Chinese, (middle) Indian, and (right) German subgroups.

Table 13 Chinese, Indian, and German Subgroup *t* Tests

Nationality	*t*	*df*	Sig. (2-tailed)	Mean difference LR/SW	95% confidence interval of difference
Chinese	−4.260	102.572	0.00	−2.46	[−3.65, −1.33]
Indian	3.762	94.688	0.00	2.18	[1.05, 3.40]
German	1.099	76.407	0.28	0.66	[−0.57, 1.96]

Note. LR = listening/reading; SW = speaking/writing.

Investigation of Profiles of TOEFL iBT® *Section Scores*

Following Bridgeman et al. (2015), we investigated a possible imbalance between listening and reading scores, on one hand, and speaking and writing scores, on the other. We calculated the difference between the listening/reading (LR) and the speaking/writing (SW) test scores and examined the distribution of these differences for the three largest nationality subgroups of Chinese ($n = 72$), Indian ($n = 67$), and German ($n = 57$) test takers. Figure 3 shows the distribution of the differences between the LR and the SW test scores for each of these groups compared to the rest of the sample.

We conducted *t* tests to check whether the differences between the LR and the SW test scores found within the subgroups differ significantly from the patterns observed for the rest of the sample (Table 13). The *t* tests showed significant differences for the Chinese and Indian subgroups compared to the rest of the sample. However, the magnitude of the differences was rather small (around 2 points). Even the extreme cases were not as pronounced as those reported by Bridgeman et al. (2015), who found differences of 16 points and more; in our sample, the largest difference of 14 points occurred only three times in the Chinese subgroup. We controlled effects on correlation and found that these score profiles do not unduly influence the results.

Furthermore, we examined differences in the final academic grades in relation to different profiles of TOEFL iBT section scores. Looking for differences in final grades given different profiles of TOEFL iBT section scores, we performed *t* tests on the differences between LR and SW scores for each of the grades (lower degree, pass, merit, and distinction; fail was discarded because it contained too few samples, $n = 4$). All *t* tests showed no evidence for a difference between grades on the profiles of the TOEFL iBT section scores. Consequently, we will not exclude any students from our analysis on the basis of their imbalanced TOEFL iBT score profiles.

Correlational Analyses of the Relation Between TOEFL iBT® *Scores and Academic Outcomes*

We now report findings of the correlational analyses with regard to the relation between the TOEFL iBT scores and the final academic grades. We expected this relation to be influenced by the following factors: (a) academic discipline, indicated by faculty; (b) requirements in the level of numeracy versus language, that is, selected departments with a focus on numeracy versus those with a focus on social sciences; (c) mother tongue, indicated by nationality; and (d) additional support in the form of pre- and in-sessional classes. In addition to correlations for the total sample, we analyzed correlations for different subgroups corresponding to these factors. Given our sample size, we had to limit the formation of subgroups to those of reasonable size. For each grouping, we reported the correlation between the overall TOEFL iBT score as well as the TOEFL iBT section scores (reading, writing, speaking, and listening) with the final academic grade.

The final grades were coded from 1 (*fail*) to 5 (*distinction*). The correlations were reported for both the uncorrected, direct correlation between the TOEFL iBT variable of interest and the final academic grade (using Spearman's rho for

Table 14 Correlations Between TOEFL iBT Scores and Academic Outcome for Total Sample

	Total sample					
TOEFL iBT	Mean	*SD*	ρ	Adj. ρ	0.5 CI adj.	*p* value adj.
Reading	26.77	2.73	0.10*	0.25*	[0.03, 0.43]	.024
Writing	25.99	2.58	0.12*	0.22*	[0.05, 0.37]	.010
Speaking	25.26	2.68	0.18**	0.30**	[0.16, 0.43]	<.001
Listening	26.71	2.77	0.22**	0.48**	[0.30, 0.61]	<.001
Overall	104.75	7.70	0.20**	0.47**	[0.28, 0.61]	<.001

Note. $N = 483$.

*Correlation significant at .05, two-tailed. **Correlation significant at .01, two-tailed.

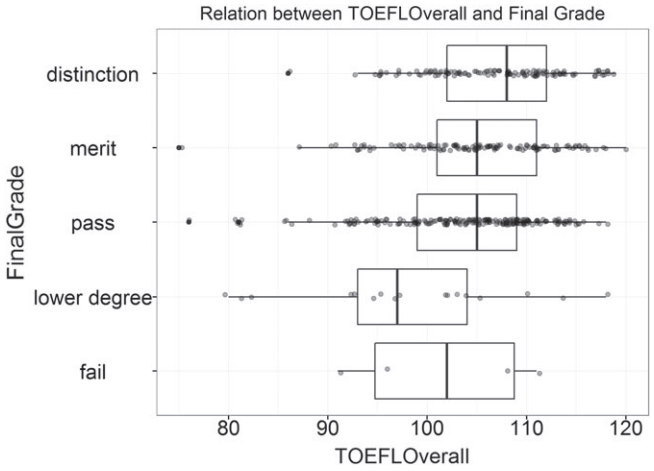

Figure 4 Relation between TOEFL iBT overall and academic outcomes.

all analyses, because the final academic grades constitute ordinal data; Bland, 2000) and the correlation estimate after correcting for the range restriction (Thorndike Case 2; Sackett & Yang, 2000). The Thorndike Case 2 correction formula is

$$r_{\text{Adj}} = \frac{\left(S_x/s_x\right) r}{\sqrt{1 + r^2\left[\left(S_x/s_x\right)^2 - 1\right]}},$$

where S_x and s_x are the standard deviations for the unrestricted and restricted populations, respectively. The confidence interval for the unadjusted correlations is calculated using the formula $\tanh(\text{atanh}(\rho) \pm 1.96/\sqrt{n-3})$ (Bonett & Wright, 2000). The confidence interval of the adjusted correlation is built by applying the Thorndike correction to the bounds of the confidence interval of the unadjusted correlation (Hunter & Schmidt, 2004). The correlations used for the analyses are always the adjusted correlations, and we report both the *p* value and the adjusted confidence intervals.

The standard deviations of the unrestricted TOEFL iBT scores were taken from the Test and Score Data Summary for TOEFL iBT Tests for 2014 report: TOEFL iBT Reading, 6.7; TOEFL iBT Writing, 5; TOEFL iBT Speaking, 4.6; TOEFL iBT Listening, 6.8; and TOEFL iBT overall, 20. Because the unrestricted standard deviations were higher than the observed (i.e., restricted) ones reported earlier, the corrected correlations are always higher than the direct correlations.

Correlations for the Total Sample

Table 14 shows the observed correlations (using Spearman's rho) and the adjusted correlations (using Thorndike Case 2) between TOEFL iBT scores and academic outcome for all students. The highest correlations are found between the listening and overall scores and the academic outcomes, followed by speaking. All correlations but the ones for reading and writing are significant at the .01 level.

Figure 4 illustrates the relation between the TOEFL iBT overall scores and the final academic grade graphically. The graph shows the distribution of the TOEFL iBT overall scores for each final academic grade. As can be seen from the

Table 15 Correlations Between TOEFL iBT and Academic Outcomes for Three Faculties

TOEFL iBT	Mean	*SD*	ρ	Adj. ρ	0.5 CI adj.	*p* value adj.
Arts						
Reading	26.56	2.46	0.29	0.64	[−0.11, 0.88]	.084
Writing	26.67	2.24	0.48**	0.77**	[0.38, 0.91]	.003
Speaking	26.25	2.53	0.36*	0.58*	[0.07, 0.82]	.029
Listening	26.75	2.57	0.31	0.65	[−0.07, 0.88]	.070
Overall	106.22	6.92	0.46**	0.83**	[0.42, 0.94]	.004
Science[b]						
Reading	25.97	3.26	0.17*	0.34*	[0.04, 0.57]	.030
Writing	25.10	2.85	0.14	0.24	[−0.03, 0.47]	.085
Speaking	24.25	2.57	0.18*	0.31*	[0.04, 0.53]	.026
Listening	25.75	3.38	0.29**	0.52**	[0.27, 0.69]	.001
Overall	101.08	9.02	0.25**	0.50**	[0.22, 0.69]	.002
Social sciences[c]						
Reading	27.21	2.32	0.04	0.10	[−0.23, 0.40]	.549
Writing	26.43	2.33	0.03	0.07	[−0.17, 0.31]	.571
Speaking	25.69	2.61	0.15**	0.26**	[0.07, 0.43]	.009
Listening	27.21	2.25	0.15*	0.41*	[0.10, 0.63]	.013
Overall	106.56	6.20	0.14*	0.41*	[0.07, 0.64]	.019

*Correlation significant at .05, two-tailed. **Correlation significant at .01, two-tailed.
[a]$n = 36$. [b]$n = 155$. [c]$n = 287$.

outliers to the left in Figure 4, students with the lowest TOEFL iBT scores managed to pass and even achieve a merit, while there is a cluster of three students with scores around 80 who received a lower degree. Students who failed were not those with low TOEFL iBT scores, indicating that there may have been other than language aspects involved.

Correlations for Subgroups by Faculty

Next, we explore correlations for the three largest faculties to examine the effect of different disciplines. We excluded Medicine ($n = 5$) due to the small sample size. Table 15 shows the results.

Interestingly, the strongest correlations show for the smallest faculty, arts, where writing, overall, and speaking scores are significantly related to the final academic grades. The science faculty shows weaker but significant correlations for all but the writing scores, while social sciences show significant correlations only for speaking, listening, and overall scores. Figure 5 gives a graphical illustration of the relations for the three faculties. The trends reported for the overall cohort can also be seen for the faculties, with a more pronounced relation for arts and science faculties.

Correlations for Selected Subgroups

Because our findings from the qualitative interview data reported in Section 6.4.3 strongly suggested differences in the role language plays in academic progress for selected disciplines (represented in our interviews), we used the interview findings to group disciplines into two groups, one of which we call *selected disciplines with a quantitative focus* (selQUANT), the other of which we call *selected disciplines with a social sciences focus* (selSOC). We selected the following majors based on the majors we had sampled in our interviews (all other majors listed in Table 8 were excluded):

- selQUANT: Business School, Mathematics Institute, Manufacturing Group, Economics, Statistics
- selSOC: Centre for Applied Linguistics, Politics and International Studies, Sociology

The correlations for the two groups of selected departments are reported in Table 16.

The correlations for departments with a quantitative focus are only significant for TOEFL iBT overall, listening, and speaking, and they mirror the strength reported for the overall sample. For the departments with a social sciences focus, the correlations are around zero. This suggests that TOEFL iBT scores are not related to academic success for this subsample, in contrast to the subsample with a quantitative focus. Figure 6 gives a graphical illustration of the relations for both subgroups.

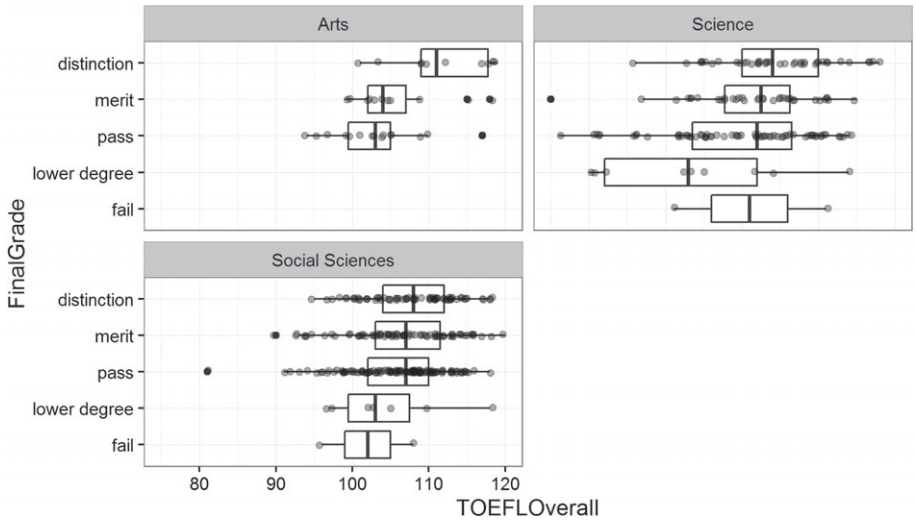

Figure 5 Relation between TOEFL iBT overall and academic outcomes for faculties.

Table 16 Correlations by Selected Disciplines With a Quantitative Focus and Selected Disciplines With a Social Sciences Focus Grouping

TOEFL iBT	Mean	*SD*	ρ	Adj. ρ	0.5 CI adj.	*p* value adj.
selQUANT						
Reading	26.80	2.80	0.05	0.12	[−0.14, 0.36]	.357
Writing	26.09	2.60	0.08	0.15	[−0.05, 0.34]	.141
Speaking	25.29	2.77	0.20**	0.32**	[0.15, 0.46]	<.001
Listening	26.79	2.83	0.21**	0.46**	[0.24, 0.62]	<.001
Overall	104.99	7.96	0.17**	0.39**	[0.15, 0.58]	.002
selSOC[b]						
Reading	26.79	2.64	−0.04	−0.10	[−0.62, 0.50]	.766
Writing	25.79	2.58	−0.07	−0.13	[−0.55, 0.36]	.612
Speaking	25.14	2.57	0.12	0.22	[−0.25, 0.58]	.356
Listening	26.19	2.61	0.05	0.13	[−0.49, 0.65]	.701
Overall	103.91	6.88	0.04	0.12	[−0.55, 0.67]	.761

Note. selQUANT = selected disciplines with a quantitative focus. selSOC = selected disciplines with a social sciences focus.
*Correlation significant at .05, two-tailed. **Correlation significant at .01, two-tailed.
[a]$n = 331$. [b]$n = 57$.

As can be seen from Figure 6, there is a more pronounced relation for the selQUANT departments, but the results have to be treated somewhat carefully due to the different sample sizes. To further investigate whether the small and nonsignificant correlation in the selected social sciences departments is due to a particular discipline, we conducted correlational analyses for each department (Centre for Applied Linguistics, $n = 17$; Politics and International Studies, $n = 32$; Sociology, $n = 8$), yet none of the correlations is significant, likely due to the small sample sizes. We will further investigate our results in light of the qualitative findings reported later.

Correlations for Subgroups by Nationality

Following Bridgeman et al. (2015), and taking into consideration the nationalities that show the largest numbers in our sample (see Table 6), we examined correlation patterns for different groups of nationalities, as these can be used as a proxy for the students' mother tongue. We now report the correlations for the three best represented nationalities in our sample, Chinese, German, and Indian, again for both observed and adjusted correlations (Table 17).

Our Chinese subgroup shows the strongest correlations, which are significant for the speaking and overall scores. For the Indian subgroup, the correlations are only significant for listening, while the German subgroup shows the smallest

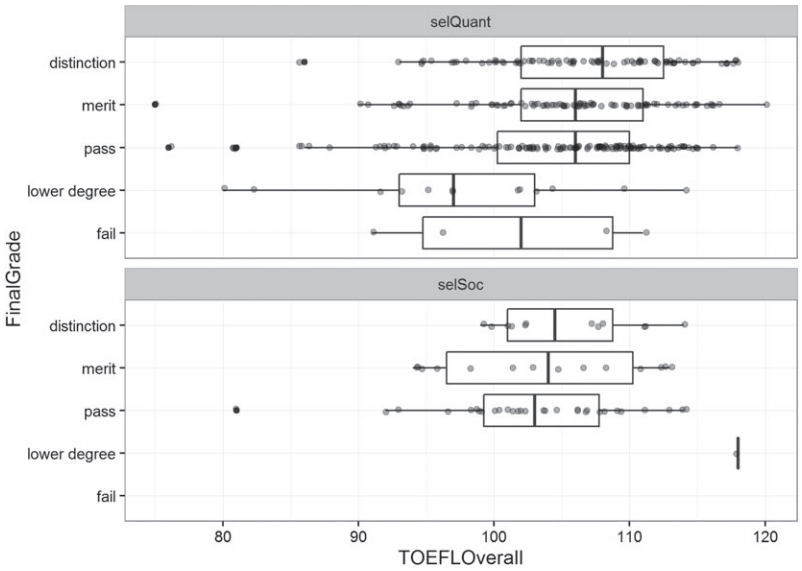

Figure 6 Relation between TOEFL iBT overall and academic outcome for selected disciplines with a quantitative focus versus selected disciplines with a social sciences focus.

Table 17 Correlations by Nationality

TOEFL iBT	Mean	*SD*	ρ	Adj. ρ	0.5 CI adj.	*p* value adj.
Chinese[a]						
Reading	27.18	2.51	0.18	0.44	[−0.14, 0.76]	.127
Writing	25.64	2.81	0.19	0.33	[−0.08, 0.62]	.109
Speaking	23.35	2.10	0.32**	0.60**	[0.21, 0.80]	.006
Listening	26.15	2.74	0.23	0.51	[0.00, 0.77]	.051
Overall	102.32	7.34	0.31**	0.66**	[0.22, 0.85]	.009
Indian[b]						
Reading	26.69	2.13	0.15	0.43	[−0.28, 0.79]	.224
Writing	26.94	2.10	0.11	0.26	[−0.30, 0.66]	.370
Speaking	26.31	2.48	0.20	0.35	[−0.08, 0.65]	.108
Listening	26.88	2.56	0.25*	0.57*	[0.03, 0.81]	.040
Overall	106.82	6.37	0.22	0.57	[−0.08, 0.84]	.077
German[c]						
Reading	26.97	2.38	−0.01	0.03	[−0.59, 0.62]	.937
Writing	26.32	2.27	0.04	0.09	[−0.45, 0.57]	.762
Speaking	27.00	2.02	0.21	0.43	[−0.13, 0.75]	.126
Listening	27.97	1.96	0.11	0.35	[−0.49, 0.80]	.434
Overall	108.35	4.86	0.14	0.49	[−0.47, 0.86]	.314

*Correlation significant at .05, two-tailed. **Correlation significant at .01, two-tailed.
[a]$n = 72$. [b]$n = 67$. [c]$n = 57$.

correlations, none of which is significant. We acknowledge that sample size and TOEFL iBT score distributions may have a certain influence on the magnitude and significance of the correlations (such as the TOEFL iBT scores for the German subgroup being relatively higher with a smaller variance, and the German subgroup being the smallest group, which could contribute to the nonsignificant results). For this reason, we corrected the correlations and the confidence intervals to account for the truncated sample and the nonnormal TOEFL iBT score distribution. Hence we would cautiously interpret that there seems to be a trend for the Chinese subgroup whereby relatively lower TOEFL iBT scores (overall and listening) show a stronger relation to the final academic grade. Across all nationalities, we cautiously conclude that the relatively higher TOEFL iBT scores show a less pronounced relation to the final academic grade. Being the largest group by nationality (see Table 6), it is worth in the case of the Chinese subgroup to further "peel the onion" (Bridgeman et al., 2015) and explore potentially hidden correlation patterns.

Table 18 Correlations Between Chinese Versus Non-Chinese Subgroups

TOEFL iBT	Mean	*SD*	ρ	Adj. ρ	0.5 CI adj.	*p* value adj.
Chinese[a]						
Reading	27.18	2.51	0.18	0.44	[−0.14, 0.76]	.127
Writing	25.64	2.81	0.19	0.33	[−0.08, 0.62]	.109
Speaking	23.35	2.10	0.32**	0.60**	[0.21, 0.80]	.006
Listening	26.15	2.74	0.23	0.51	[0.00, 0.77]	.051
Overall	102.32	7.34	0.31**	0.66**	[0.22, 0.85]	.009
Non-Chinese[b]						
Reading	26.70	2.76	0.10*	0.25*	[0.02, 0.44]	.034
Writing	26.05	2.55	0.10*	0.19*	[0.00, 0.36]	.046
Speaking	25.60	2.63	0.13*	0.22*	[0.05, 0.37]	.011
Listening	26.81	2.77	0.20**	0.46**	[0.26, 0.60]	<.001
Overall	105.17	7.71	0.16**	0.39**	[0.17, 0.56]	.001

*Correlation significant at .05, two-tailed. **Correlation significant at .01, two-tailed.
[a] $n = 72$. [b] $n = 411$.

Table 19 Correlations Between Chinese Subgroups in Business Schosol

	TOEFL iBT Reading	TOEFL iBT Writing	TOEFL iBT Speaking	TOEFL iBT Listening	TOEFL iBT overall
Whole school[a]	0.07	−0.12	0.36*	0.46	0.54
	[−0.44, 0.53]	[−0.46, 0.27]	[0.09, 0.58]	[−0.04, 0.73]	[−0.04, 0.8]
Non-Chinese[b]	0.11	−0.24	0.33	0.51	0.44
	[−0.44, 0.58]	[−0.56, 0.17]	[0.00, 0.58]	[−0.02, 0.78]	[−0.25, 0.78]
Chinese[c]	0.12	0.56	0.37	0.04	0.73
	[−0.86, 0.89]	[−0.58, 0.91]	[−0.59, 0.85]	[−0.79, 0.80]	[−0.75, 0.96]

*Correlation significant at .05, two-tailed. **Correlation significant at .01, two-tailed.
[a] $n = 153$. [b] $n = 127$. [c] $n = 26$.

Correlations for the Chinese Subgroup

Following Bridgeman et al.'s (2015) results, and taking into account the largest group by nationality in our sample, which also showed the strongest correlations, we now examine correlation patterns for the Chinese subgroup, also within different departments. First, we compare correlations for the Chinese and non-Chinese subgroups (Table 18).

It is interesting to note that the non-Chinese subgroup shows a distinctively different pattern from the Chinese, Indian, and German subgroups we have examined, with all correlations being significant. The concept of significance, however, has to be interpreted in relation to sample size, as indicated previously. The *p* values in the non-Chinese group may in part be driven by the larger sample size. Hence the effects of nationality on correlation patterns have to be interpreted cautiously, as we have indicated. Nevertheless, what we can state is a pronounced trend within the Chinese subsample for TOEFL iBT overall and speaking scores (which were relatively lower in this subgroup as compared to the non-Chinese group) to show a closer relationship with the final academic grade than any of the TOEFL iBT scores show in the non-Chinese group.

We further examined the correlation patterns for Chinese students within the Business School (the department with the largest number of students) and within the selQUANT (see earlier), because this grouping showed substantial correlations (as opposed to the selSOC). For space reasons, we now report only the adjusted correlation rho (and adjusted confidence intervals in brackets; Table 19).

Here, while correlations for the whole school and the non-Chinese subgroup show similar patterns in size, none of the correlations (apart from speaking for the whole school) become significant. Within the selQUANT, however, a different picture arises (Table 20).

Correlations are significant for the TOEFL iBT overall, listening, and speaking scores for all selQUANT students and the non-Chinese students, albeit with slightly different degrees of strength. It is interesting that here the Chinese subgroup shows the strongest correlations, which are significant for overall, writing, and speaking scores. Also interestingly, the reading scores are not significant for any of the subgroups here, which may in part be driven by the fact that the TOEFL iBT reading scores show a ceiling effect and the smallest variance in all subgroups.

Table 20 Correlations Between Chinese Subgroups in Selected Disciplines With a Quantitative Focus

	TOEFL iBT Reading	TOEFL iBT Writing	TOEFL iBT Speaking	TOEFL iBT Listening	TOEFL iBT overall
All[a]	0.12	0.15	0.32**	0.46**	0.39**
	[−0.14, 0.36]	[−0.05, 0.34 9	[0.15, 0.46]	[0.24, 0.62]	[0.15, 0.58]
Non-Chinese[b]	0.11	0.07	0.25**	0.43**	0.29*
	[−0.16, 0.36]	[−0.16, 0.28]	[0.06, 0.42]	[0.20, 0.61]	[0.01, 0.51]
Chinese[c]	0.52	0.60*	0.58*	0.56	0.78**
	[−0.21, 0.83]	[0.16, 0.82]	[0.01, 0.83]	[−0.06, 0.83]	[0.34, 0.91]

*Correlation significant at .05, two-tailed. **Correlation significant at .01, two-tailed.
[a]$n = 331$. [b]$n = 288$. [c]$n = 43$.

Table 21 TOEFL iBT Score Differences for Students With/Without Additional Language Support

	TOEFL iBT Reading[a]	TOEFL iBT Writing[b]	TOEFL iBT Speaking[c]	TOEFL iBT Listening[d]	TOEFL iBT overall[e]
Students with support[f]	25.00 (3.58)	24.32 (3.05)	23.38 (2.47)	24.28 (3.43)	96.99 (9.09)
Students with no support[g]	27.09 (2.41)	26.29 (2.38)	25.60 (2.58)	27.15 (2.39)	106.15 (6.53)

Note. Standard deviations are in parentheses.
[a]$t(85.38) = -4.822$, $p < .001$. [b]$t(89.71) = -5.261$, $p < .001$. [c]$t(103.82) = -7.077$, $p < .001$. [d]$t(86.29) = -6.897$, $p < .001$. [e]$t(87.13) = -8.290$, $p < .001$. [f]$n = 74$. [g]$n = 409$.

Table 22 Final Academic Grades for Students With/Without Additional Language Support

Support	Fail	Lower degree	Pass	Merit	Distinction
Language support[a]	0 (0%)	5 (6.8%)	36 (48.6%)	21 (28.4%)	12 (16.2%)
Students with no support[b]	4 (1%)	12 (2.9%)	175 (42.8%)	106 (25.9%)	112 (27.4%)

[a]$n = 74$. [b]$n = 409$.

Students With Additional Language Support

As outlined previously, 74 students in our sample received extra support in the form of pre- and in-sessional classes. We aim to measure the impact of these classes on the relation between TOEFL iBT scores and students' academic outcomes. Students attending these classes are mostly students who had a lower TOEFL iBT score than the average student population; for example, they were accepted on the academic course with slightly lower TOEFL iBT scores than the threshold set by the academic department, but on the condition that they attend presessional classes. Some students also joined voluntarily.

We first compare TOEFL iBT scores and final academic grades across these two groups, using a *t* test to test for statistically significant differences. Table 21 shows the two groups' TOEFL iBT mean scores, standard deviations (in parentheses), and the *t* test results.

Levene's test revealed that we cannot assume equal variances. Hence the variances were estimated separately using the Welch formula. The *t* test results are significant at the .001 level and indicate that the two groups show significant differences in all mean test scores.

We now compare the two groups' distributions of final academic grades. Table 22 cross-tabulates absolute numbers and percentages (within the two groups).

Chi-square tests of independence were performed to examine the relation between language support and the final academic grade. The chi-square tests showed no significant difference between the two groups, $\chi^2(4) = 7.01$, $p = .136$ (Pearson chi-square). These results indicate that we cannot conclude that there are differences in the final academic grades between students who received language support (and entered with significantly lower TOEFL iBT scores) and those who did not receive language support. The box plot in Figure 7 illustrates the differing distributions of the TOEFL overall scores for these two subgroups.

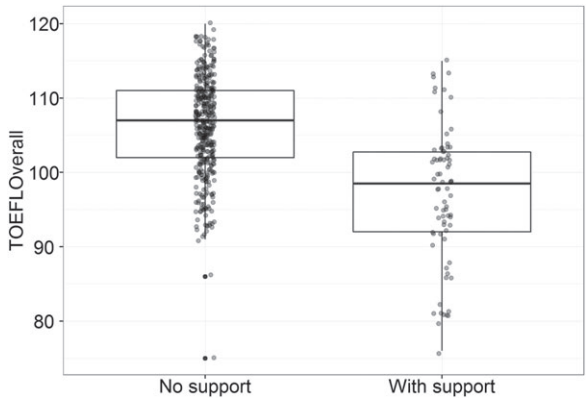

Figure 7 Box plot of TOEFL iBT overall scores for students with/without additional language support.

Table 23 TOEFL iBT Subgroups by Faculty

Faculty	Overall bottom 25%	Overall top 25%
Arts[a]	102	110
Science[b]	95	108
Social sciences[c]	103	111

[a] $n = 36$. [b] $n = 155$. [c] $n = 287$.

While the TOEFL iBT scores of students who received language support are lower on average, as was to be expected and as shown by the *t* test results, we know from the chi-square test results that these students do not show differences in their final academic results compared to students with no language support, who enter with significantly higher TOEFL iBT scores. This seems to indicate that language support classes are successful insofar that students who enter with lower TOEFL iBT test scores and receive language support classes do not seem to be disadvantaged with regard to their final academic grades.

Predictive Power of TOEFL iBT Scores on Final Academic Grades

We now explore the predictive power of the TOEFL iBT test scores on the final academic grades to answer RQ2. First, we used so-called expectancy graphs as an efficient way of summarizing the data, as suggested by Cho and Bridgeman (2012), before we examined the TOEFL iBT scores as predictors of final academic grades by using an ordered linear regression (Agresti, 2002).

Expectancy Graphs

Cho and Bridgeman (2012) suggested cross-tabulating TOEFL iBT scores and final academic grades in expectancy graphs to display the predictive validity in terms of students in one TOEFL iBT score subgroup belonging to one of the five final academic grades. Following their approach, we divided the TOEFL iBT overall score into three subgroups: the bottom 25% range, the middle 50% range, and the top 25% range. We did this for the three faculties (we excluded Medicine due to the small sample size), with the cut scores for the TOEFL iBT overall score per faculty as shown in Table 23.

Figure 8 displays the three expectancy graphs for the three largest faculties. The three vertical bands within each graph represent the TOEFL iBT score subgroups, while the color-coded bars represent the percentage of students within a certain TOEFL iBT subgroup achieving one of the five possible final grades. Owing to space limitations within the graphs, we display the percentages for all five final grades in Table 24.

Table 24 shows the respective percentages for the three groups of bottom 25%, middle 50%, and top 25% of TOEFL iBT scores differentiated by their academic outcome, separate for the three faculties.

What can be seen from the expectancy graphs and the percentages in Table 24 is a trend for students in the higher TOEFL iBT score subgroups to achieve a higher academic grade (merit or distinction), while a higher percentage of

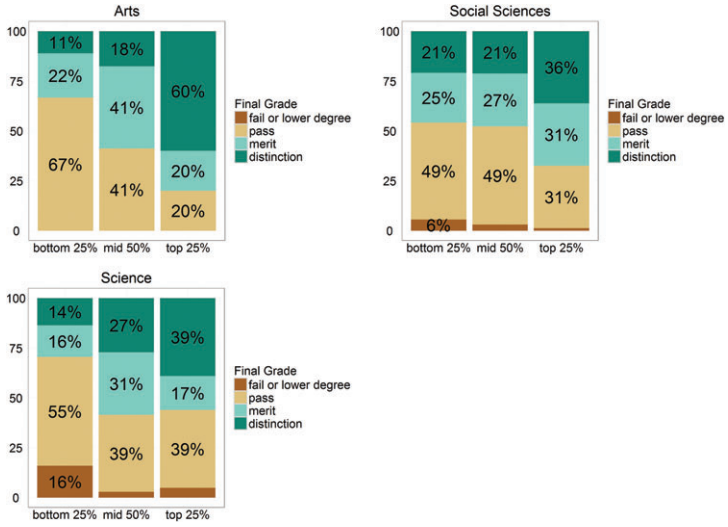

Figure 8 Expectancy graphs by faculty.

Table 24 Students (in %) Within TOEFL iBT Subgroups Achieving a Certain Final Academic Grade

Final academic grade	Bottom 25%	Middle 50%	Top 25%
Arts			
Pass	66.67	41.18	20.00
Merit	22.22	41.18	20.00
Distinction	11.11	17.65	60.00
Science			
Fail	2.27	0.00	2.44
Lower degree	13.64	2.86	2.44
Pass	54.55	38.57	39.02
Merit	15.91	31.43	17.07
Distinction	13.64	27.14	39.02
Social sciences			
Fail	1.39	0.76	0.00
Lower degree	4.17	2.27	1.20
Pass	48.61	49.24	31.33
Merit	25.00	26.52	31.33
Distinction	20.83	21.21	36.14

students in the bottom and mid-range TOEFL iBT subgroups are awarded a *pass*. Lower academic degrees and fails are observed as more likely in the bottom and are rarely seen in the mid-range TOEFL iBT subgroup. This trend is most pronounced among the arts faculty, but it can also be observed in sciences, where 13.6% of the "low" TOEFL iBT students have this outcome, while only 2.4% of the "high" TOEFL iBT students have this outcome. That is, more than 5 times as many "low" TOEFL students have this poor outcome compared to "high" TOEFL students. The trend is also clear at the top end of the scale, with about 30% of the "low" TOEFL iBT students earning merit or distinction compared to 56% of the top TOEFL iBT students. The trend is equally clear in the social sciences, with fewer than half of the "low" TOEFL iBT students earning merit or distinction compared to more than two-thirds of the "high" TOEFL iBT students.

This trend is even more pronounced when grouping students by the aforementioned selSOC and selQUANT; for space reasons, we do not include the expectancy graphs here.

Though we found low to moderate correlations with various levels of significance for different subgroups, the expectancy graphs give a somewhat clearer picture of the relation between certain TOEFL iBT score bands and certain academic grades, with a clear trend for students in higher TOEFL iBT bands to achieve higher academic grades, while students who fail or receive a lower academic grade are most likely to be found in the bottom TOEFL iBT score band.

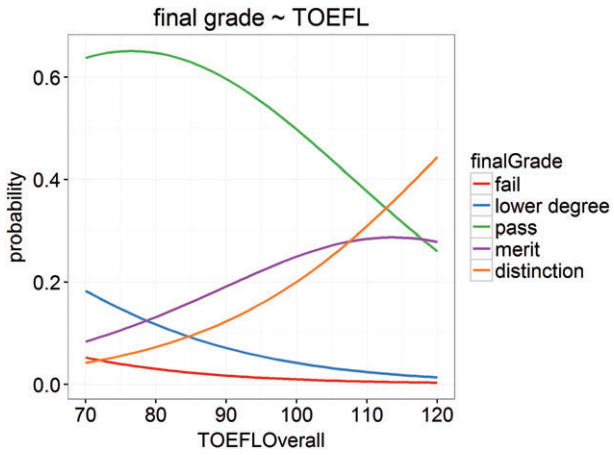

Figure 9 Ordered logistic regression model for TOEFL iBT overall score.

Regression Analyses

We now examine the TOEFL scores as predictors of final academic grades by using an OLR (Agresti, 2002) to model the predictive relation between TOEFL iBT scores and final academic grades. In this approach, the TOEFL iBT scores were used as independent variables to predict the academic grades as dependent variables. We first examined the predictive relation between the TOEFL iBT overall score and the TOEFL iBT section scores and students' final academic grades before we took selected variables (students' nationality, academic disciplines, additional language support) as predictors into our model.

Predictive Power of TOEFL iBT Scores

We first fit a regression model with the TOEFL iBT overall score only. TOEFL iBT overall scores and final academic grades are moderately but significantly correlated on the whole population ($\rho = .20$; adjusted $\rho = .43$; see earlier). From this model (final grade ~ TOEFL iBT overall), we took the TOEFL iBT overall score as the independent variable and derived the probabilities of achieving a certain final academic grade, as depicted in Figure 9 (there are five possible grades, as indicated by the five lines in the figure).

As displayed in Figure 9, *pass* has the highest probability for TOEFL iBT overall scores from 70 up to 112. To put things into perspective, the lowest overall TOEFL iBT score in our sample is 75, for which the model predicts a pass with a probability of .62. For TOEFL iBT scores higher than 112, the most probable outcome becomes distinction.

Using this model, we predicted the probabilities for achieving each of the five possible final academic grades from the TOEFL iBT overall score only: For each student, the model estimated the probabilities of achieving a certain academic grade on the basis of the student's TOEFL iBT overall score. That final grade estimated as the most probable one was then compared with the actual final grade obtained by the student. On the basis of this comparison, the model's accuracy could be determined. Using the logistic regression on TOEFL iBT overall, we got a model accuracy of 44.72%. This means that our model predicts the correct final academic grade in 44.72% of all cases, based on a calculation of the percentage of correctly predicted cases. To put things in perspective, a linear regression[12] yields a model accuracy of 30.85%. Compared to the most simple regression model using only the most frequent grade (pass) as the predicted value, which yields a model accuracy of $211/483 = 43.69\%$, our model achieves only a slightly better accuracy.

We have to concede that we have only a small number of scores at the lower end of our TOEFL iBT score spectrum, making predictions less accurate at the lower end. Hence we will treat estimations at the lower end rather like explorations.

We now turn to exploring cumulative probabilities, that is, the likelihood of getting *at least* a pass rather than *exactly* a pass. We assumed that this information is also of value for test score users. Hence Figure 10 depicts the probabilities of getting at least a pass, as compared to getting a fail or lower degree.

We found that even for the lowest TOEFL iBT overall score of 75 in our sample, the probability of getting at least a pass is above 75%, constantly increasing with increasing TOEFL iBT scores. Conversely, the probability of getting a fail or lower degree is just below 25% for the lowest TOEFL iBT scores, steadily decreasing with increasing TOEFL iBT scores.

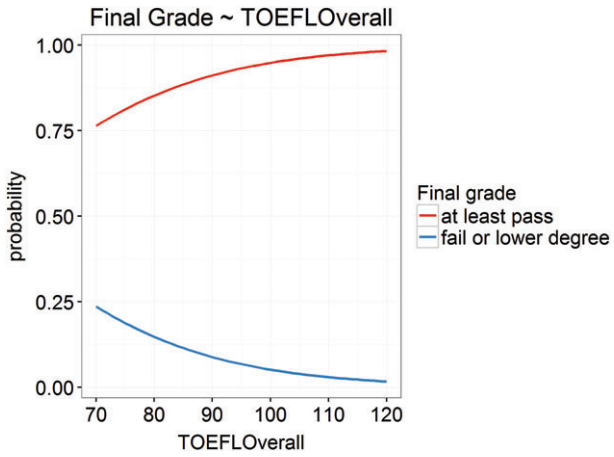

Figure 10 Probabilities for "at least a pass."

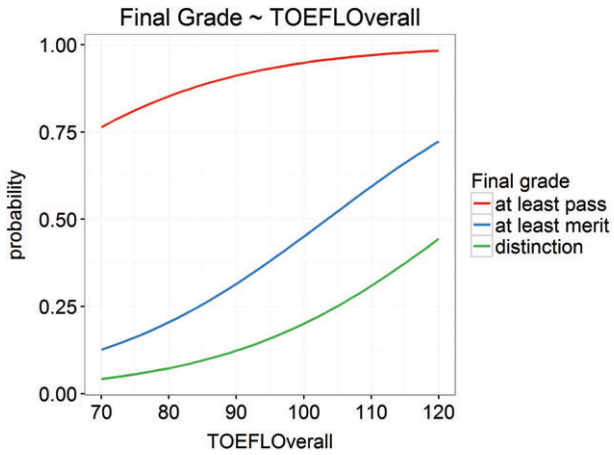

Figure 11 Probabilities for "at least a merit or a distinction."

If we use this model to predict at least a pass, the accuracy is 95.65%. Next, we examine the cumulative probabilities of getting at least a merit or a distinction, as depicted by the blue and green lines in Figure 11.

We see an increase in probabilities for merit and distinction particularly for the higher scores, as was to be expected. Yet the probabilities for merit or distinction never surpass the probability of getting a pass.

Next, we examined separate models for the four TOEFL iBT section scores. The four models yielded almost identical results; while the actual probabilities to obtain a certain final academic grade were slightly different, the predictions were very similar. For section scores between 13 and 28, the highest probability is to achieve a pass; the lowest section score in our sample is 16. The model accuracies, or in other words, their predictive power, are as follows: TOEFL iBT Writing, 43.69; TOEFL iBT Reading, 43.69; TOEFL iBT Listening, 43.89; and TOEFL iBT Speaking, 44.31. Hence the section scores do not offer greater predictive power over the TOEFL iBT overall score alone. This could partly be related to the previously examined correlations among TOEFL iBT overall and section scores.

The correlation analyses and expectancy graphs have brought to light that certain subgroups, grouped by certain variables, relate differently to the final academic grade. Hence, in the following sections, we include the two variables that yielded promising relational patterns, that is, faculties and nationalities, as extra factors in the regression model.

Predictive Power of TOEFL iBT by Faculty

Taking the TOEFL iBT overall score and the faculty into the OLR, we modeled the final academic grade as an interaction of TOEFL iBT and Faculty: final grade ~ TOEFL iBT Overall × Faculty. We removed Medicine as it only represents five

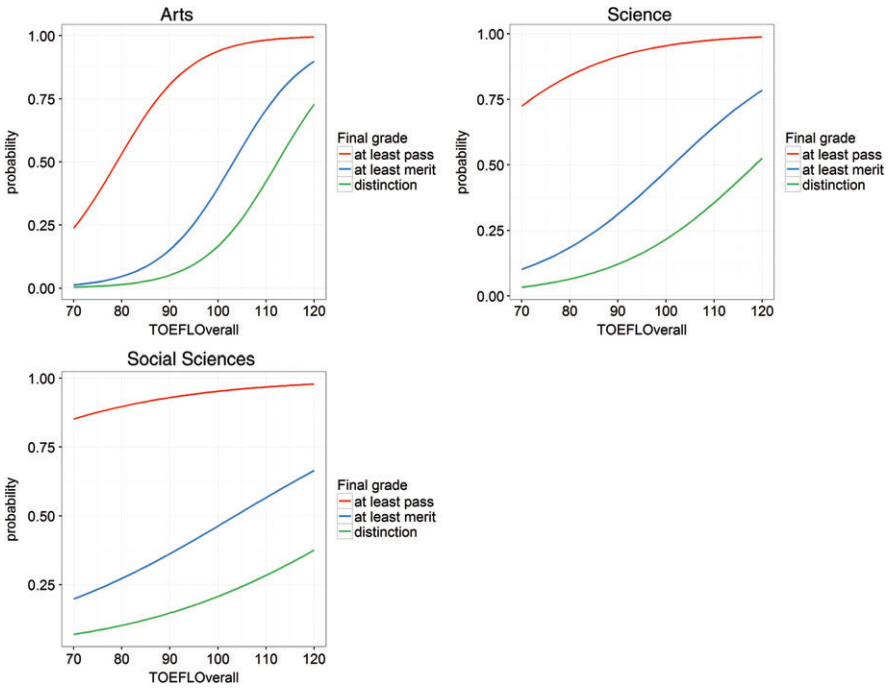

Figure 12 OLR model for TOEFL iBT overall scores by three faculties.

students. The three graphs in Figure 12 correspond with the three faculties. As before, the red line indicates the cumulative probabilities for "at least a pass," the blue line the probabilities for "at least a merit," and the green line the probabilities for a distinction.

Interestingly, the results for science and social sciences show a remarkable similarity to the overall TOEFL iBT model (see Figure 9) for all students. Only the arts faculty shows a pattern differing more pronouncedly from the reference model, starting off at the lower end with a much lower probability for achieving at least a pass, and higher probabilities for achieving at least merit or distinction at the upper end of the TOEFL iBT score range. When looking for the TOEFL iBT overall score where at least a pass gets more likely than achieving a lower degree or failing, for the arts faculty, this transition point is located at a TOEFL iBT overall score of 79, while for the other faculties, even for TOEFL iBT overall scores as low as 70, our model predicts the most likely outcome of at least a pass. Yet we have to acknowledge the low number of students in the arts faculty as well as the lower accuracies at the lower end of the TOEFL iBT score range, hence a lower accuracy of prediction in these regions, and thus be careful not to overinterpret these predictions.

This model predicts the correct final academic grade with 45.61% accuracy (again calculated as explained earlier, i.e., computing the percentage of correctly predicted final academic grades). A naive model (using only the most frequent final grade of pass) achieves 43.51% accuracy. Compared to the reference model reported earlier (TOEFL iBT overall only, 44.56% accuracy, for students from these three faculties), the model taking faculties into account achieves only a slightly higher accuracy.

We also examined regression models separately for the two largest departments, Business School and Manufacturing, with both models predicting the transition from pass to distinction at a TOEFL iBT overall score of 114. The models differ in their prediction of receiving a pass for a TOEFL iBT overall score of 70: The probability for students in the Business School is slightly higher than in the reference model and slightly lower for those in Manufacturing. The predictive power for the Business School model with 42.5% is lower than that of the reference model, while it is slightly higher for Manufacturing, with 48.6%

Predictive Power of TOEFL iBT by Nationality

In analogy to the model taking faculty into account, we now take TOEFL iBT overall score and the three best represented nationalities (Chinese, $n = 72$; Indian, $n = 67$; German, $n = 57$) into the OLR, modeling the final academic grade as a

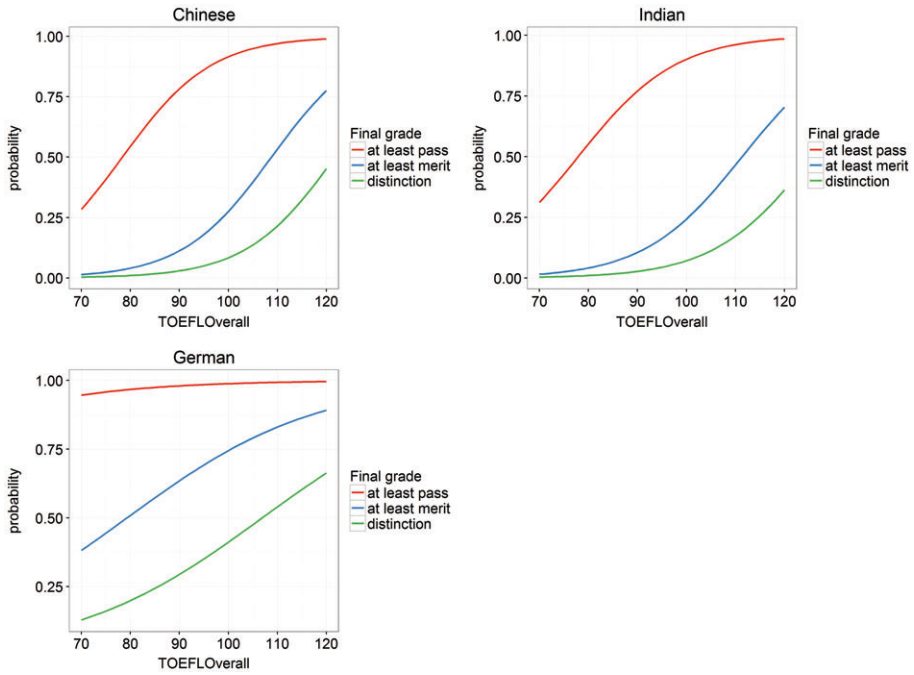

Figure 13 OLR model for TOEFL iBT overall scores by nationality.

regression on TOEFL iBT and nationality: final grade ~ TOEFL iBT Overall × Nationality. The three graphs in Figure 13 correspond to the three selected nationalities; the lines in the graphs again refer to the probabilities of getting at least a pass, at least a merit, or a distinction.

The model taking nationality into account yields similar results for the Chinese and Indian populations, starting off with a lower probability for achieving at least a pass at the lower end of TOEFL iBT scores compared to the German subgroup, which, even for the lowest TOEFL iBT scores, has a very high probability of achieving at least a pass. Bearing test score users in mind, it may be of interest for them to look at the transition points for fail/lower degree and at least a pass, that is, to examine at which TOEFL iBT overall score a pass becomes more likely than a lower degree or fail. While for the German students, at least a pass always has the highest probability, the transition point for the Chinese and Indian students is located at a TOEFL iBT overall score of 78. Again, we have to concede that we have lower accuracies at the lower end of the TOEFL iBT score range and hence have to be careful to not overinterpret the predictions here.

The model predicts the correct final grade with 54.59% accuracy. Compared to a naive model (only predicting the most frequent final grade of pass), which yields 43.88% accuracy for this subgroup of students, and compared to the reference model (TOEFL iBT overall only), which yields 46.94% accuracy for this subgroup of students, taking nationality into the prediction increases the predictive value of the TOEFL iBT overall score considerably.

Predictive Power of TOEFL iBT and Additional Language Support

Next, we took additional language support into the OLR to examine potential effects of language support on academic outcome, while accounting for the fact that students in this group came in with significantly lower TOEFL iBT scores. We modeled the final academic grade as an interaction of TOEFL iBT and extra language courses. Not surprisingly, the model predictions for the group who did not receive additional support are identical to Model 1, while for the group of students who came in with lower TOEFL iBT scores and hence received additional language support, we find a slight drop in the probability for receiving a pass at the lower end of the TOEFL iBT scores, at 57% for TOEFL iBT overall of 65 and 62% for TOEFL iBT overall of 70 (the lowest TOEFL iBT overall score in our sample is 75). Looking at the probabilities of achieving at least a pass, students with language support have a 65% probability even as low as a hypothetical TOEFL iBT overall score of 65, while students in the group with no language support have a probability of 72% at TOEFL iBT overall of 65. The model predicts a slightly lower transition point from pass to distinction at TOEFL iBT overall of 110 as compared to 113 for the rest of the students in this model. The model accuracy, or its predictive power with 44.31%,

is slightly lower than that of the reference model (TOEFL iBT overall, 44.72%), so that we do not gain predictive power when taking additional language support into the model. This is not to say that additional language support would not be useful; rather, our results support the conclusion that language support is justified and effective, because the students who come in with lower TOEFL iBT scores have comparable (albeit slightly lower) probabilities of getting at least a pass even in the lowest TOEFL iBT score range.

Summary

We examined the relation between TOEFL iBT test scores and final academic grades and found differing strengths and significances for different subgroups. For the total sample, we found weak but significant correlations for all TOEFL iBT scores. We then grouped different disciplines, based on the existing faculties. Here the strongest correlations show for the smallest faculty, arts, where writing, overall, and speaking scores are significantly related to final academic grades. The science faculty shows weaker but significant correlations for all but the writing scores, whereas social sciences shows weak significant correlations only for speaking, listening, and overall scores. We then grouped students, based on the qualitative interview findings, into selQUANT and selSOC; when comparing these two groupings, only selQUANT showed weak correlations for speaking, listening, and overall TOEFL iBT scores, whereas there were no significant correlations for selSOC; however, it has to be noted that the selSOC grouping contained few students. Grouping students by the three largest nationalities, Chinese, Indian, and German, also yielded interesting results: The Chinese subgroup, being the largest in number, showed the strongest correlations, which were significant for the speaking and overall scores. For the Indian subgroup, the correlations were only significant for listening, whereas the German subgroup showed the smallest correlations, none of which was significant; one has to bear in mind that the German group contained few students.

Interestingly, like Bridgeman et al. (2015), we found that Chinese students exhibit a slightly different TOEFL iBT profile than the rest of the population. However, in our sample, it was not as pronounced as what Bridgeman et al. reported so that we did not exclude any students on these grounds.

When examining the effect of additional language support classes, we needed to take into account the fact that students with additional language support tended to have lower TOEFL iBT scores than students without language support; a *t* test confirmed significant differences in mean TOEFL iBT scores (overall and for all section scores) between the two groups. However, when comparing differences in final academic grades between the two groups, a chi-square test (*p* value at .136) indicated that we cannot conclude that there are differences in the final academic grades between the two groups. As a matter of fact, students who took language support classes tended to come with lower TOEFL iBT scores but nevertheless showed no measurable differences in their final academic grades in comparison to students who did not attend language support classes.

Next, we investigated the potential of the TOEFL iBT scores with regard to predicting students' final academic grades. The expectancy graphs showed a tendency for students in higher TOEFL iBT score subgroups to achieve higher academic grades, whereas students who failed or achieved lower grades were more likely to be found in lower TOEFL iBT subgroups. This trend is most pronounced in the arts faculty, as was to be expected from the correlation results, but it can also be observed in sciences and social sciences. This trend becomes even clearer when grouping students by the aforementioned selQUANT versus selSOC. In addition, we conducted OLR analyses to model the predictive relation between TOEFL iBT scores, the additional predictors of faculty and nationality, and final academic grade. Using the TOEFL iBT overall score alone, we could improve over a naive model by 1% to a model accuracy of 44.72%. Using the four different TOEFL iBT section scores as predictors did not add anything over the model that used only the TOEFL iBT overall score. Adding faculty as a predictor had only little effect. Looking separately at the two biggest departments also did not add much accuracy, nor did adding additional language support as a predictor, whereas adding nationality as a predictor improved the model accuracy over the TOEFL iBT overall model by 10%. Hence, for our data set, TOEFL iBT overall scores and nationality are the strongest predictors, yielding a predictive power of 54.59%. When looking at the predictive power of a model predicting at least a pass, the TOEFL iBT overall model yields over 95% accuracy.

Interestingly, all models predict pass as the most probable outcome for the lowest possible TOEFL iBT scores (for scores as low as TOEFL iBT overall of 70 and section scores as low as 15), and the transition from pass to distinction in the region of 112 to 113, apart from the model that takes nationality into account, where there is a distinctively lower transition point from merit to distinction for the German subgroup at TOEFL iBT overall 94 and higher transition points for the Chinese

(116) and Indian (119) groups; for the latter two groups, we found transition points for getting at least a pass at TOEFL iBT overall of 78. Looking at the lower end of TOEFL iBT scores, we found that students who received additional language support still had a 57% probability of achieving a pass and a 65% probability of achieving at least a pass for a TOEFL iBT overall score of 65, with a transition point of distinction getting more likely as a final outcome than pass or merit at a TOEFL iBT overall score of 110.

Findings From Strand 2

We first present findings separately for the students and the tutors and separately for the different instruments we used within these samples. This approach allows insights from the different perspectives, which will be brought together to answer our research questions in the Interpretation and Discussion of Findings section. Before we present findings from the questionnaires and interviews, we describe the respective student and tutor samples and their background characteristics in detail.

Details on Strand 2 Student Sample

To give the reader an overall idea of our total student sample in Strand 2 ($n = 48$), we provide background characteristics collected in the two questionnaires ($n = 23$) and the interviews ($n = 25$) on students' age when taking up their studies, gender, first language, the department they studied with, and their final academic grades (from Central Registry for interview students; Table 25) and TOEFL iBT scores (self-reported in questionnaires; for interview students who provided their university ID, confirmed by Central Registry data; Table 26).

To sum up, the Strand 2 student sample encompasses roughly 50% men and women, with an average age of 26 years. Our sample is characterized by a large variety of first languages and departments. The students in general came with a high average TOEFL iBT score (105.6). With regard to these characteristics, the Strand 2 sample is a representative sample of our reference sample from Strand 1 (see the Data Summary, Descriptive Statistics subsection). With regard to the final academic grade, we cannot claim representativeness, because this information was only available for 25 out of the 48 Strand 2 participants. Those 25 students whom we could identify in the Central Registry data (via the student IDs collected in the interviews) successfully finished their studies, the majority with a merit or distinction.

Details on Strand 2 Tutor Sample

To give an overall idea of the tutor sample participating in Strand 2, in Table 27, we provide background details as reported in the questionnaire and interviews on the tutors' gender, first language, and department. Tutors' length of experience ranged from 0.3 to 26.4 years, with a mean 11.55 and a standard deviation of 8.994 years.

To sum up, the tutor sample is characterized by a slight majority of men and by a variety of first languages, with the majority of tutors (72.4%) reporting English as their first language. Our tutor sample covers nine EAP tutors (15.5%) located in the Centre for Applied Linguistics; the academic lecturers work in the departments where our students are located; the tutors work in a range of roles and positions and have an average of 11.5 years of work experience, thus representing the necessary diversity to allow insights from all relevant perspectives.

Findings From the Questionnaires

We now present the findings from the questionnaires separately for the students and the tutors. The results here serve as background and will be taken up again in the Interpretation and Discussion of Findings section, when we draw on the results gained from our different instruments to answer our research questions.

Student Questionnaire 1

The first student questionnaire ($n = 31$ students) targeted students' preparation for the TOEFL iBT test, for their academic studies, and for daily life in the United Kingdom; students' perceptions of their preparedness for the linguistic demands during their study abroad period; students' perceptions of how well the TOEFL iBT test prepared them for these linguistic requirements; and how well the TOEFL iBT scores reflect students' linguistic skills.

Table 25 Strand 2 Student Sample Demographics

Demographic	Frequency	Percentage
Gender		
Female	23	47.9
Male	25	52.1
First language		
Arabic	2	4.2
Bahasa Indonesia	1	2.1
Bengali	1	2.1
Chinese	4	8.3
Dutch	1	2.1
English	1	2.1
French	3	6.3
German	2	4.2
Greek	1	2.1
Hindi	4	8.3
Italian	10	20.8
Japanese	2	4.2
Korean	1	2.1
Mandarin	2	4.2
Mongolian	1	2.1
Portuguese	1	2.1
Serbian	1	2.1
Spanish	7	14.6
Swedish	1	2.1
Thai	1	2.1
Not mentioned	1	2.1
Department		
Applied Linguistics	2	4.2
Bio-Economy	1	2.1
Chemistry	2	4.2
Complexity Science	3	6.3
Economics	4	8.3
German Studies	1	2.1
Italian Studies	2	4.2
Mathematics	1	2.1
Politics/International Studies	6	12.5
Physics	1	2.1
School of Engineering	5	10.4
Sociology	1	2.1
Statistics	4	8.3
Business School	5	10.4
Manufacturing Group	8	16.7
Not mentioned	2	4.2
Final academic grade		
n/a[a]	6	12.5
Pass	7	14.6
Merit	5	10.4
Distinction	7	14.6
Missing (questionnaires only)	23	47.9
Total	48	100.0

Note. $N = 45$. Student sample mean age was 25.82 years, min. 21 years, max. 33 years, $SD = 3.625$.
[a]Six students were on part-time or PhD programs not leading to a final grade after 1 year.

TOEFL iBT Test Preparation

We asked students how they prepared for the TOEFL iBT and how useful they rated the different means of preparation that are offered on the official TOEFL iBT Web pages on a scale from 1 (*not useful*) to 5 (*very useful*). Table 28 shows the results.

The second column *n* indicates how many students made use of a specific preparation means, with the majority of our sample using the free test preparation materials offered by ETS. It is noteworthy that only five students attended a

Table 26 Strand 2 Student Sample TOEFL iBT Scores

TOEFL iBT	*N*	Min.	Max.	Mean	*SD*
Overall	47	86	117	105.64	7.230
Listening	46	17	30	27.30	2.988
Speaking	46	22	30	24.93	2.195
Reading	46	21	30	27.41	2.409
Writing	46	17	30	25.63	2.977

Table 27 Strand 2 Tutor Sample Demographics

	Frequency	Percentage
Gender		
Male	25	43.1
Female	32	55.2
Not stated	1	1.7
First language(s)		
Brazilian	1	1.7
Bulgarian	1	1.7
English	42	72.4
English; French	1	1.7
Finnish	1	1.7
German	2	3.4
Hindi	1	1.7
Russian	1	1.7
Spanish	2	3.4
Turkish	1	1.7
Vietnamese	1	1.7
Not stated	4	6.9
Department		
Applied Linguistics[a]	20	34.5
Politics/International Studies	6	10.3
School of Engineering	8	13.8
Sociology	2	3.4
Statistics	4	6.9
Business School	9	15.5
Manufacturing Group	9	15.5

Note. $N = 31$.

[a]Note that all EAP tutors are part of the Centre for Applied Linguistics, which also hosts academic lecturers.

Table 28 Usefulness of Preparation Means for TOEFL iBT

SQ1.4 TOEFL iBT test preparation	*n*	Min.	Max.	Mean	*SD*
SQ1 4.1 Priced test preparation	9	4	5	4.56	.527
SQ1 4.2 Skill-building tools	3	3	5	3.67	1.155
SQ1 4.3 Free test preparation	21	2	5	3.86	.964
SQ1 4.4 Attended test prep course	5	4	5	4.40	.548
SQ1 4.5 Test prep other	14	3	5	4.57	.646

Note. The numbering refers to the item numbers in the questionnaire.

preparation course. With regard to the usefulness of the different means, students perceived the priced test preparation as most useful (4.56) and the skill-building tools and free materials as useful. Fourteen students reported a range of other preparations, such as watching TV shows or YouTube videos and using TOEFL iBT practice books and CDs for self-study, which they generally rated as useful to very useful.

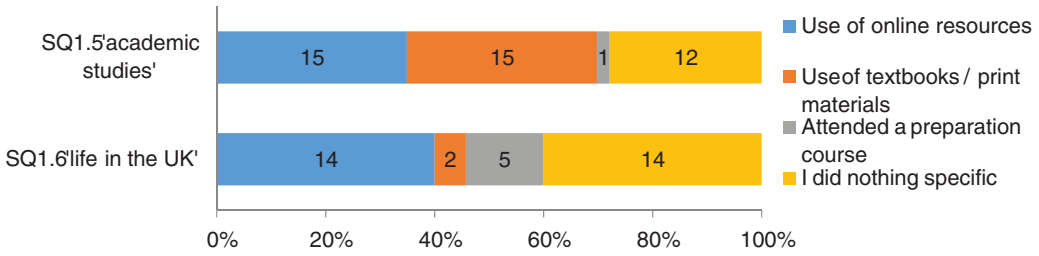

Figure 14 Preparation for academic studies and for life in the United Kingdom.

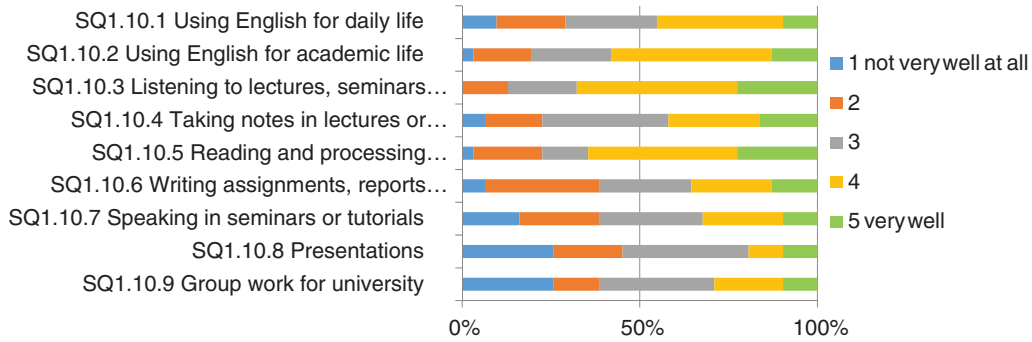

Figure 15 Q1 students' perception of TOEFL iBT as indicator for preparedness.

Preparation for Academic Studies and for Life in the United Kingdom

We then asked students how they prepared for their academic and social lives in the United Kingdom. Students could tick a combination of several options and state other means of preparation. The results are shown in Figure 14.

The majority of students made use of online resources and textbooks to prepare for their visit to the United Kingdom, with a very small number attending preparation courses. A high number of students did nothing specific to prepare either for their academic studies or for life in the United Kingdom, which seems fairly reasonable when considering the fairly high English-language ability levels represented in our sample.

Linguistic Preparedness

Students were asked in a yes/no form whether they felt that their English was good enough to cope with the linguistic requirements of their academic studies and of daily life, as well as whether they expected to improve their English during their stay in the United Kingdom (SQ1.6/7/8). An overwhelming majority of 96.8% (30 students) said their English was good enough to cope with the academic demands, 87.1% (27 students) felt their English was good enough to cope with everyday English, and 90.3% (28 students) expected their English to improve. Overall, these answers indicate that students felt well prepared while being aware that their English had the potential for improvement.

TOEFL iBT as Indicator of Preparedness

We asked students whether they thought that the TOEFL iBT test prepared them well for the linguistic requirements at university. Thirty-one students answered this question and rated their perception of TOEFL iBT on a scale from 1 (*not well at all*) to 5 (*very well*), referring to the range of language aspects in Figure 15.

Students reported that the TOEFL iBT test prepared them rather well for the receptive skills of listening (average 3.77) and reading (average 3.61), while they felt less well prepared for the productive skills of speaking (average 2.87), giving presentations (average 2.58), and group work (average 2.74). Interestingly, students felt the TOEFL iBT test prepared them better for using English in academic life (average 3.48) than in everyday life (average 3.16). Given the coverage of TOEFL iBT, these ratings are in line with what would be expected.

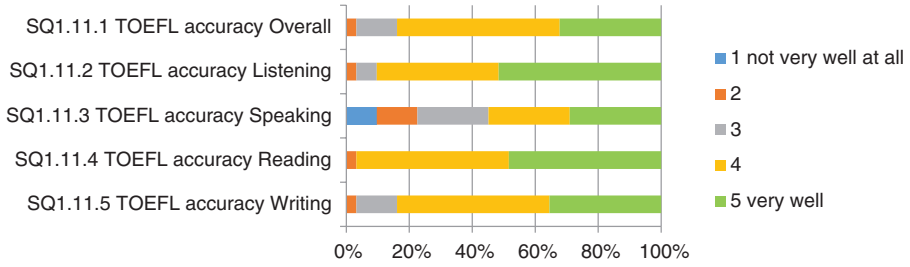

Figure 16 Q1 students' perceptions of TOEFL iBT as indicator for language skills.

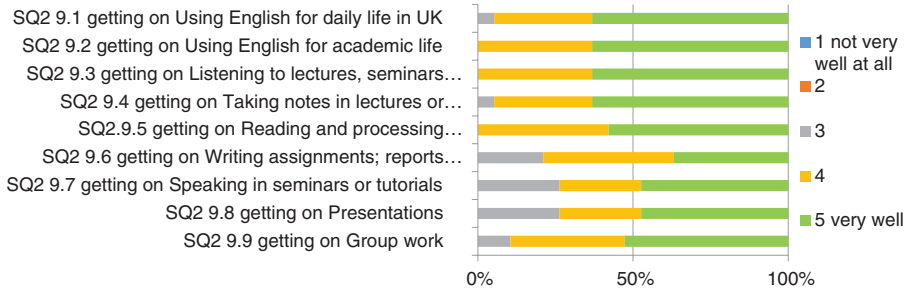

Figure 17 Q2 students' perceptions of how well they got on linguistically.

TOEFL iBT as Indicator of Language Skills

We also asked students for their perceptions of how well the TOEFL iBT test scores reflect their linguistic skills overall and how well the four subskills reported by the TOEFL iBT profile reflect students' actual subskills, again using a 5-point scale ranging from 1 (*not well at all*) to 5 (*very well*). Figure 16 is a graph of the results.

The 31 students who answered this question generally rated TOEFL iBT as an accurate measure of their language skills, with the TOEFL iBT reading section score the most highly rated as an accurate reflection of their skill (average 4.42), while the speaking section score was perceived as somewhat less accurate (average 3.52).

Student Questionnaire 2

The second student questionnaire (Q2) was filled in by 19 students, 8 of whom had also filled in the first questionnaire, Q1. The items in Q2 mirrored those in Q1 in as many aspects as possible; Q2 entailed items targeting the themes of students' perceptions of coping with the linguistic demands and improving their English as well as students' perceptions of TOEFL iBT as an indicator of preparedness and as an accurate measure of their linguistic skills.

Linguistic Preparedness

Mirroring the items SQ1.6/7 from Q1, students in Q2 were asked whether their English was good enough to cope with the linguistic requirements of their academic studies and of daily life (SQ2.6/7). In Q2, all 19 students stated that their English was indeed good enough to cope with the academic requirements, and 89.5% stated that their English was good enough to cope with daily life. Results from Q2 support students' positive assumptions in Q1 about their linguistic preparedness.

In Q2, we wanted to get a nuanced picture of how students thought they had coped with different linguistic demands during the past year. We used the same list of linguistic aspects as in item SQ1.10, this time, however, with a focus on students' perceptions of how well they got on in these different areas, using a 5-point scale (Figure 17).

The 19 students in Q2 reported a high level of coping with the linguistic demands, with no average ratings below 4, indicating that they got on very well. This speaks indeed for students having been well prepared for the language demands imposed by their studies and by everyday life. Interestingly, students in Q2 rated the level with which they were coping even more highly (average 4.44) than students in Q1 rated their perceived level of preparedness by the TOEFL iBT test (average 3.17; see item SQ1.10). It seems that students were better prepared than they had initially thought.

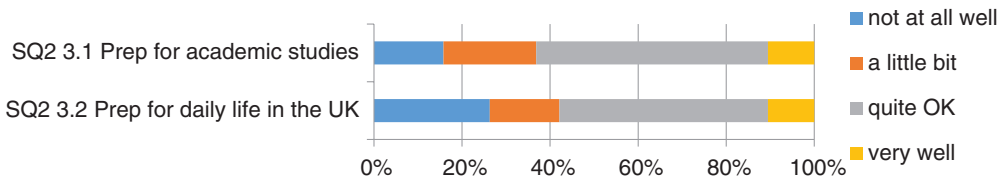

Figure 18 Q2 students' perceptions of TOEFL iBT as indicator for preparedness.

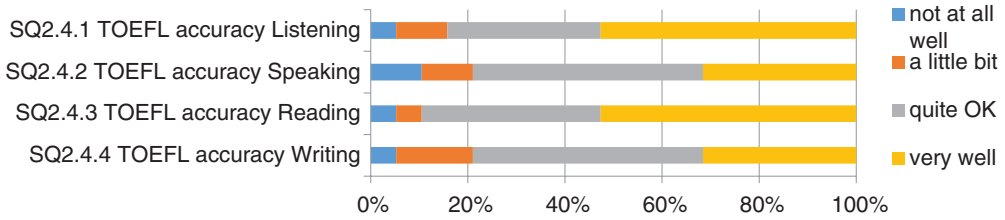

Figure 19 Q2 students' perceptions of TOEFL iBT as indicator of language skills.

Improvement of Language Skills

The picture is more varied with regard to students' perceptions of having improved their language skills during their stay in the United Kingdom (SQ2.8, mirroring SQ1.8). Students were asked in Q2 to rate this aspect on a scale from 1 (*no improvement*) to 5 (*improved a lot*), resulting in an average of 3.11 (*SD* .81, range 2–4). Hence, while students in Q1 were initially expecting to improve their English, students in Q2 reported only a moderate level of improvement. We tried to capture some of the potential reasons for (non-)improvement by asking whether students exploited tuition in any way to improve their English (SQ2.15), which 94.7% denied; only one student received tuition. When asked whether they had tried to actively improve their English over the year, only five students (26.3%) said that they had done so. This low level of exploitation of English-language support might somewhat account for students' perceptions that their English did not improve considerably. Alternatively, one could assume that students' initial levels of language proficiency were so high that only little improvement was to be realistically expected. We will use interview data to shed further light on students' varying reasons for (not) exploiting language support.

TOEFL iBT as Indicator of Preparedness

We asked students at the end of the year whether they thought that the TOEFL iBT test prepared them well for the linguistic requirements of their academic studies and of daily life. Students rated their perceptions of TOEFL iBT on a 4-point scale ranging from 1 (*not at all well*) to 4 (*very well*), referring to the two aspects of academic language and language used in daily life (Figure 18).

In Q2, the majority of students perceived that the TOEFL iBT test prepared them quite well for the linguistic demands during their studies, with the perceptions of the TOEFL iBT preparing students for daily life slightly lower. Compared to students' answers in Q1 (SQ1.10.2 and SQ1.10.1) at the beginning of the year, this perception has not changed dramatically. However, because we only have a small number of students who filled in both questionnaires ($n = 8$), we need to be careful not to overinterpret this perception.

TOEFL iBT as Indicator of Language Skills

In analogy to Q1, we asked students at the end of the year for their perceptions of how well the TOEFL iBT test scores reflected their linguistic skills overall and how well the four subskills reported by the TOEFL iBT profile reflected students' actual subskills, again using a 5-point scale ranging from 1 (*not well at all*) to 5 (*very well*). Figure 19 shows the results.

Students in Q2 perceived TOEFL iBT as indicating their language skills quite OK to very well, with LR being rated relatively higher and SW being rated lower. This trend is similar to the one reported for Q1.

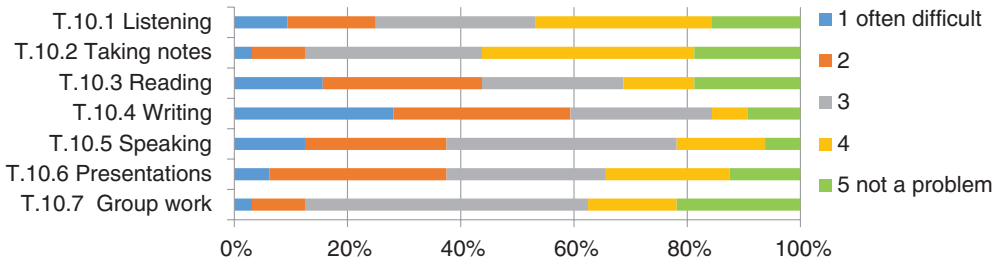

Figure 20 Tutors' perceptions of international students' linguistic preparedness. The numbering refers to the item numbers in the questionnaire.

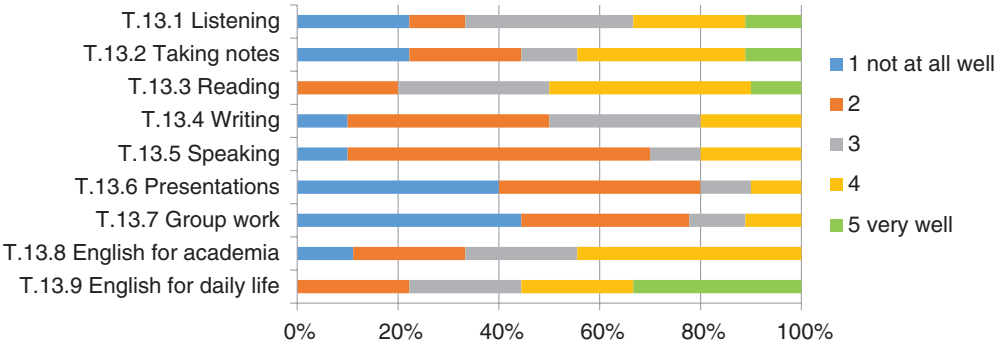

Figure 21 Tutors' perceptions of TOEFL iBT as indicator of students' linguistic preparedness.

Tutor Questionnaire

The tutor questionnaire ($n = 32$ tutors) encompassed a number of items targeting tutors' perceptions of how well international students cope with the linguistic requirements of their studies; how well the TOEFL iBT test prepares students for these linguistic requirements; how well the TOEFL iBT scores reflect students' linguistic skills; how well the TOEFL iBT test predicts academic success; the tutors' perceptions of the usefulness of TOEFL iBT test reports; and tutors' familiarity with the Common European Framework of Reference, a language proficiency framework that is gaining importance in the HE setting.

Linguistic Preparedness

In the questionnaire, we asked for tutors' perceptions of how international students, that is, students who are not British and whose first language is not English, in general get on with the English-language requirements and in which areas tutors perceive these students may be struggling. Tutors rated the aspects shown in Figure 20 on a scale from 1 (*often difficult*) to 5 (*not a problem*).

It seems that tutors perceive students to be getting on satisfactorily in the areas of note taking (average rating of 3.59) and group work (average 3.44), followed by listening (average 3.28). According to tutors' perceptions, students appear to get on less but still acceptably well with the productive skills of writing (average 2.38) and speaking (average 2.78).

TOEFL iBT as Indicator of Preparedness

Tutors were asked whether they thought that the TOEFL iBT test prepared students well for the linguistic requirements at university. Tutors rated their perception of TOEFL iBT on a scale from 1 (*not well at all*) to 5 (*very well*), again referring to the same language aspects as in the items above (Figure 21).

It is noteworthy that 22 out of 32 tutors stated that they were not familiar enough with the TOEFL iBT test to answer this question. We did provide a link to the TOEFL iBT test embedded within the questionnaire, so that tutors had the opportunity to familiarize themselves with the test. On the basis of the answers of those tutors who felt comfortable enough to answer, the TOEFL iBT test seems to prepare students quite well for the English they need in daily life (average

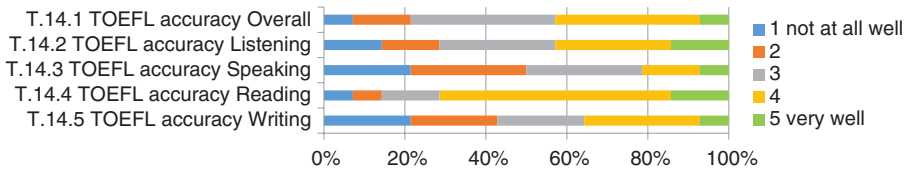

Figure 22 Tutors' perceptions of TOEFL iBT as indicator of students' language skills.

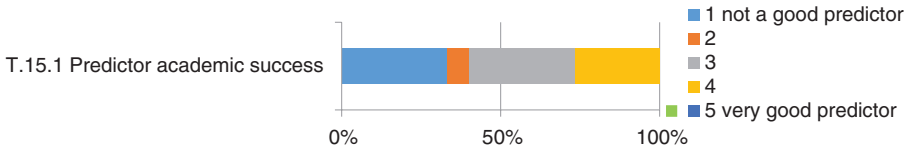

Figure 23 Tutors' perceptions of TOEFL iBT as predictor of academic success.

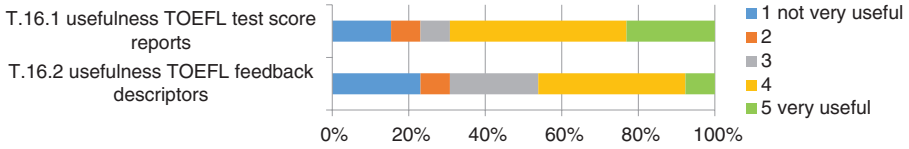

Figure 24 Tutors' perceptions of usefulness of TOEFL iBT test reports.

rating of 3.67) and for academic reading (average 3.40). It appears that the areas where TOEFL iBT is perceived to prepare less well are found in group work (average 1.89) and presentations (average 1.90). Given the small number of tutors these results are based on, we have to treat them with caution.

TOEFL iBT as Indicator of Language Skills

We also asked tutors for their perceptions of how well the TOEFL iBT test scores reflect students' linguistic skills overall and how well the four subskills reported by the TOEFL iBT section scores reflect students' actual subskills, again using a 5-point scale ranging from 1 (*not well at all*) to 5 (*very well*). (See Figure 22.)

Here, 18 tutors stated they were not familiar enough with the TOEFL iBT test to answer. Based on 14 answers, it appears that tutors perceived the TOEFL iBT reading section score to be quite an accurate reflection of students' academic reading skills (3.64), whereas the speaking section score was regarded as the least accurate (2.57) in reflecting students' actual speaking skills.

TOEFL iBT as Predictor of Academic Success

Next, tutors were asked for their perceptions of TOEFL iBT as an adequate predictor of academic success, again on a 5-point scale ranging from 1 (*not a good predictor*) to 5 (*very good predictor*; Figure 23).

Here 17 tutors said they were not familiar enough with the test to answer. The remaining 15 tutors rated TOEFL iBT as a fairly good predictor (average 2.53) of academic success, yet with a relatively high standard deviation (1.25), indicating a range of opinions. No one rated TOEFL iBT as a very good predictor.

Usefulness of Test Reports

Tutors were also asked how useful they found the TOEFL iBT test score reports and the accompanying qualitative TOEFL iBT feedback descriptors, again on a scale from 1 (*not very useful*) to 5 (*very useful*). The questionnaire contained a link to a score report example for tutors to familiarize themselves if needed (Figure 24).

Although 19 tutors said they were not familiar enough with the test to answer, the majority of the remaining 13 tutors rated the TOEFL iBT score reports as useful (average 3.54) and the descriptors as a little less useful (average 3.00). Here again, however, we have a range of opinions.

Appropriateness of Entry Requirements

Finally, we asked tutors whether they regarded the existing English-language entry requirements of their departments as too low (1), appropriate (2), or too high (3). The majority, 23 tutors, stated that the requirements were appropriate; nine tutors regarded them as too low; and none of the tutors regarded the entry requirements as too high. This indicates that tutors generally thought entry requirements were set at an appropriate, perhaps slightly too low, level.

Comparing Students' and Tutors' Perceptions as Reported in the Questionnaires

We will now compare the results from the student and tutor questionnaires with regard to students' and tutors' perceptions of students' linguistic preparedness and their perceptions of TOEFL iBT as an indicator of preparedness and as an indicator of students' actual language skills.

Linguistic Preparedness

All questionnaire data suggest that students are well enough prepared to cope with the linguistic demands of academic studies in the United Kingdom. Students at the beginning and end of the year reported with an overwhelming majority that their English was good enough to cope with the language demands of both academic and daily life. More specifically, students at the end of the year rated their level of coping with different linguistic demands even higher than the tutors did, and both groups agreed that students coped better in the areas of reading, listening, and note taking, while the ratings in both groups were somewhat lower for the productive areas of writing and speaking.

TOEFL iBT as Indicator of Preparedness

Data from all three questionnaires showed that participants rated the power of the TOEFL iBT test to indicate students' preparedness at a satisfactory to high level. With regard to different linguistic aspects, TOEFL iBT's indicative power was ranked by all groups in the same order, that is, highest for the receptive skills, lower for the productive skills, and lowest for group work and giving presentations. Students in both questionnaires thought TOEFL iBT had prepared them better for the linguistic demands of academic life, whereas tutors perceived the test to prepare students better for language demands in daily life.

TOEFL iBT as Indicator of Language Skills

All participants regarded TOEFL iBT as an accurate measure of students' language skills, with the student group at the beginning of the year giving the highest ratings. This could partly be due to the fact that they recently had taken the test. Interestingly, all groups rated TOEFL iBT highest in its accuracy to measure reading, followed by listening, writing, and speaking. The tendency that TOEFL iBT seems more closely related to the receptive skills can be observed both for its power to prepare students and for its potential to accurately indicate students' language skills. We used interview data to shed more light on students' and tutors' perceptions of possible reasons and explanations for this trend.

Findings From the Interviews

We present our findings separately for the three main research themes, and within these, separately for the students and their tutors, before comparing the two perspectives for triangulation. The results are taken up again in the Interpretation and Discussion of Findings section, when we draw on the results gained from our different instruments to answer our research questions.

A detailed overview of the 25 student participants and their tutors is presented in Appendix K, showing the students' participation points, tutors interviewed, departments, TOEFL iBT scores, final academic grades, and an indicator of students' "preparedness" as perceived at the end of their studies. For the latter, we built the mean for item SQ2.9 (see earlier: Student Questionnaire 2, Linguistic Preparedness; 5-point scale for nine subitems), which was used in both Interview 3 and Q2; we drew on Q2 for cases where the student had not participated in Interview 3.

Table 29 Coding Scheme for Theme 1/RQ3

NVivo parent node	NVivo child nodes	No. of sources	No. of references
1. Linguistic preparedness; getting on with English	1. Well prepared	109	962
	2. Struggling, challenges	96	613
	3. Adjustment in the beginning	21	100
	4. Improvement over time	68	281
	5. Little or no improvement over time	24	57
	6. Changing skill needs/demands	16	28
	7. Feeling prepared for dissertation stage	20	42
2. How well TOEFL prepared linguistically for studies	1. Test prepared well	39	150
	2. Test did not prepare well	32	81
	3. Limited preparation	21	46
3. How well TOEFL prepared linguistically for social life	1. Test prepared well	14	25
	2. Test did not prepare well	27	54
	3. Limited preparation	10	16
5. TOEFL as indicator of language skills	1. Effective indicator	35	57
	2. Not an effective indicator	24	45
6 TOEFL as indicator of academic grades/success/performance	1. Effective indicator	31	76
	2. Not an effective indicator	10	17
	3. Expectations	19	28
7. Preparation for TOEFL	1. Preparation for TOEFL	31	108
	2. Did not prepare	7	12
	3. Prep differently for test	15	22
8. Preparation for academic studies	1. Actual prep for academic studies	14	27
	2. Prep differently for academic studies	21	33
9. Preparation for social life	1. Actual prep for social life	8	12
	2. Prep differently for social life	17	32
10. SELT requirements		40	108

Note. SELT = Secure English Language Test.

It is interesting to note that students with relatively lower TOEFL iBT scores and indicators of preparedness, such as S036 or S039, still managed to receive a distinction, the highest academic grade, and no student failed or received a lower degree. This could be due to a self-selection effect of our Strand 2 participants, but given the low fail rates reported over 3 years in Strand 1, our Strand 2 sample seems to reflect the trends reported in Strand 1 at the selected university.

Preparedness

To analyze students' and their tutors' perceptions of students' linguistic preparedness for academic studies, we drew on Coding Theme 1/RQ3: preparedness and perception of TOEFL iBT. Table 29 gives an overview of the selected codes we analyzed here as well as a numerical summary of coding statistics on the number of sources and, within the sources, the number of references that contain the respective codes. This serves to illustrate the number of documents/interviews as well as the number of times a certain utterance was coded under the respective codes, to illustrate the magnitude and, to a certain degree, the importance of the different codes.

Students' Perceptions of Their Linguistic Preparedness and Progress

Overall, students indicated that they were "getting on well" during the first stage of the year. Students felt they were able to actively participate in academic culture. Their ability to participate in lectures and academic group work was primarily attributed to their prior use of English in their home academic environment. Many students felt they understood English with ease, as in their home countries, they studied the language from childhood. The data show that students believed they were taught the essential skills necessary to participate in both academic culture and social life, as illustrated in the following interview excerpts:

> I think that I have no difficulty with using English in daily life or using English in academic life in general, because we start from children studying English. So, for example, opening a bank account or finding accommodation, we

> study this expression over and over again in high school and primary school. So, we . . . I mean, I started studying English, probably at 8 years old, so at the beginning of primary school. (S002, Interview 1)

> The reading and listening, I have always been good at it because I am very used to it. In my country all my textbooks were in English and I watched a lot of TV that was in English so that part I don't have any trouble. (S013, Interview 1)

When discussing the skills required to succeed in their academic classes, students felt they were getting on very well in the lectures. In particular, listening and taking notes were mentioned by the vast majority of students as areas in which they were able to cope well, whereas group work and writing were often designated as the more difficult tasks they faced. Students attributed their ability to digest the lecture material to the formality of the English used in presenting the material. One student discussed how, in his home country, he had "stud[ied] English very formally, so I think it's easier for me to handle the English during the classes" (S031, Interview 1). Unsurprisingly, one of the challenges for students in listening to lectures was deciphering the range of accents encountered in the classroom.

Yet, this language hurdle was perceived to be a low threat to students' success, as it was mitigated by the lecturers' teaching style (speaking slowly and explaining specialist language), and most students stated their strongest skill was listening, whereas one of their weaker skills was comprehending academic texts. Students felt the specialist vocabulary of their field was used in unfamiliar constructs within academic texts, challenging their ability to understand the material necessary to progress with their course work. By the third interview, most students had overcome this obstacle, as they had used the material frequently for their assessments. However, students were still struggling with writing at the end of the academic year. Students from the Centre for Applied Linguistics and the Department of Politics and International Studies were required to write frequently for their assessments, whereas economics and manufacturing students were assessed less by writing and more by empirical methods. Yet, students from both groups expressed concerns about writing capability, stating they "did not think that it had improved that much" (S006, Interview 3) and that writing was the "single area in which there could still be improvement" (S005, Interview 3).

Another key challenge arose in group work, where students needed to comprehend a wide range of English spoken within the groups. Unlike the English spoken by lecturers, students found the range of English spoken by their peers challenging. Interestingly, most students were able to work in groups due to their freedom to choose their working groups. Students found the group work sessions to be useful in building their English-speaking abilities. Students reported they did not "feel nervous at all when. .. communicating with [other international students]" (S032) as they were able to "build on each other" (S033), and although they did not share a native language, they "understand each other quite well" (S019). It was apparent that most students felt most comfortable working in mixed-nationality groups.

When invited to evaluate their experiences getting on in social situations, students were all faced with similar social concerns. The most apparent concern of students focused on their ability to carry out necessary social interactions, such as finding accommodation, opening a bank account, or speaking in everyday settings. Although most had learned the phrases to participate in these social situations, many still expressed apprehensiveness in carrying out these essential tasks. One of the first challenges facing students was opening a bank account; yet even with their apprehensiveness, students generally found they were able to open a bank account, as "[the bank tellers] were really patient" (S032, Interview 1). Students reported they had used tools in their own language to simplify tasks, such as using online forums in their native language to find accommodation or using resources in their language to make sense of transportation links.

At the start of the year, two-thirds of the students stated they struggled to some degree conversing in everyday English. During the first interview stage, one student (S004) commented on his frustrations in communicating with his British housemate and in often having to ask him to repeat phrases. By the third interview stage, the student had adjusted to his housemate's accent and indicated feeling comfortable understanding and communicating with most people; yet he still struggled with "Northern English" accents due to not only the accent but also the fast pace of speech. This student's experience was not unique. Many of the students who were challenged in social situations expressed frustration with the colloquial terms used in everyday English, the speed at which someone spoke, or the strength of the individual's accent, all of which were not present to the same degree in their classes. At the end of the year, several students stated their discontent with their lack of improvement using British English, even though they indicated improving to some degree. One student stated that international students may have struggled due to their propensity to "stick with" those who speak the same native language, stating even though he "pictured our life here with all our English friends,. .. there are not really that many opportunities for us foreigners to speak with the English" (S032, Interview 3). Yet, most students believed their

speaking abilities had improved to some degree over the year, as everyday interactions provided them with opportunities to improve, which, according to the students, would not have been possible if they had been studying in another country.

Lecturers' Perceptions of Students' Preparedness

The lecturers' perceptions of students' academic preparedness were primarily positive. Lecturers reported students were generally well prepared for their academic studies. Most lecturers believed their tutees were engaged with the academic material of their courses and, although they may have been slightly shy, were willing to participate during group activities and listened intently to their tutors. However, in contrast to the students, lecturers perceived lectures to be a fundamental challenge for students, as they suggested "concentrat[ing] on understanding the language as well as the content, [was] going to be hard, and easier to get lost" (T36). For instance, one student (S017) had stated he "understood almost everything," whereas his lecturer (T36) perceived the student to be struggling, as the student would "have a slightly confused look on his face" and "there were bits he [S017] missed." Given that S017 did not achieve high marks on his final academic grades, this discrepancy may be an indicator of this student's lack of awareness of his academic skill set.

Both academic lecturers and EAP tutors found the areas where their tutees struggled were within presentations and writing more than listening or speaking informally. Several of the lecturers commented that students struggled with presenting due to a perceived lack of confidence in the classroom. Although most lecturers believed their students could communicate well in English, they found their students were not normally forthcoming to address groups. As the year progressed, lecturers recognized improvements in students' confidence. Few lecturers believed students were unable to participate in their course work due to a lack of confidence at the third stage of research.

The second fundamental challenge to student success, identified by lecturers, focused on student writing. Lecturers reported mixed reviews of student writing. Although most students were able to "process academic text well" (T14, Interview 1), a number of students struggled to produce well-written work at the start of the year. Lecturers and EAP tutors did highlight the shortcomings in their students' grasp of the grammatical structure of their writings. Grammatical errors often included issues with sentence structure, such as inverted nouns and verbs. EAP tutors/lecturers observed students were most challenged, and made more mistakes, when they were under pressure, whether that be in classroom writing situations or early assessments. Over the course of the year, only a few students did not improve in this regard, while most of the students did make progress, learning to write in the accepted academic structure necessary to their fields.

When asked to discuss their perceptions of students' ability to use English for everyday and social purposes, EAP tutors/lecturers generally perceived their tutees to be getting on well. When EAP tutors'/lecturers' statements were compared to student statements, EAP tutors/lecturers did not appear to be knowledgeable regarding students' personal issues, such as difficulty in understanding spoken discourse in some social or public situations. This is not surprising, given that most EAP tutors/lecturers were not responsible for student pastoral care. EAP tutors/lecturers were concerned that any initial communication challenges faced (such as establishing a bank account) might delay students' progress in settling down at the university. Interviews from students, however, indicated that this was not an issue.

TOEFL iBT as Preparation for Linguistic Requirements During Study Abroad

Students generally felt that TOEFL iBT prepared them well for the linguistic requirements of academic life, a little less so for daily life; this view was reinforced over the course of the year. In particular, listening and taking notes were mentioned by the vast majority of students as areas for which TOEFL iBT prepared them very well, followed by reading and writing: "To listen to teachers and to take notes. .. I think it prepared me well because I really have found something that I had prepared and done during the TOEFL test in my lectures," as S004 explained. With regard to writing, and more so speaking, students had mixed views about the test's helpfulness in preparing them. S037, for example, stated in the first interview, "For speaking section, I don't think there's that much connection with TOEFL. Maybe it does help me in terms of daily life. But for lectures and seminars and academic studies, I don't think TOEFL speaking has helped me much," whereas S034 explained in the third interview that "it helped me. It helped me perform under pressure because I had to speak to a computer." The main issues raised by students were that academic writing differed from the writing part in TOEFL iBT in style, length, and conventions, while for speaking, it was the computer-delivered mode that many students regarded as not very helpful, because they felt under pressure and could not interact with another human being. With regard to daily life, several students mentioned that the test did not prepare them for the different accents they encountered in the

United Kingdom. Interestingly, the fact that random topics are used in TOEFL iBT — commented on by several students as a limiting factor if the topic was unfamiliar — was regarded by one student (S004) in hindsight as helpful in preparing him for dealing with unfamiliar aspects during his studies. Other than for this aspect, students' views were noticeably stable across the year.

Interestingly, none of the EAP tutors or lecturers could comment on TOEFL iBT's potential to prepare students because they did not feel familiar enough with the test content and format.

Some students mentioned that their English was already good enough and hence they regarded the test as a measurement tool for certifying admission to the university. Yet other students perceived that it was the actual test preparation that facilitated their preparation for the language requirements found at university, which is why we turn to this aspect now.

Preparing for TOEFL iBT

Students reported a variety of means to prepare for the test, with none clearly emerging as a winner: They attended language courses; took test preparation courses; used ETS materials, including the practice test; searched the Web for tips; took TOEFL (and other tests) repeatedly; and used a range of other strategies, such as listening to radio, watching TV, watching movies and YouTube videos, taking notes while listening, practicing speaking by Skyping with friends or recording themselves, reading books and novels, and practicing writing by applying tips found on the Internet.

Interestingly, two students felt they were unable to prepare for the speaking section, mainly because they were required to speak to a computer. A minority of five students said they did not prepare for the test, with two nevertheless mentioning that they used the practice test and ETS materials, while one student mentioned his English was good enough so he did not need to prepare. One student (S004) did recognize the link between not having prepared for the writing part and getting a lower score in writing: "I did not actually train for the writing part of the TOEFL and I think this is where I scored least."

Looking back over the year, 12 students stated that they would not do anything differently in preparing for TOEFL iBT, two mentioned they would listen more to radio or TV, and one student would take more practice tests.

TOEFL iBT as Indicator for Language Skills

Students in general found TOEFL iBT a fairly good indicator of their language skills, particularly for listening, reading, and writing; they had mixed views about speaking, with a slight majority of students regarding the speaking score as not very well reflecting their actual speaking skills. They attributed this mainly to the computer-delivered mode, the time pressure, and unfamiliar topics. It is noteworthy that over the year, students' views on the indicative power of TOEFL iBT seemed to become more elaborate, presumably because students then had the experience of applying their language skills, thus being in a better position to judge how well the test scores reflected their actual skills. S039 stated in the second interview, "My worst score was in listening. I still struggle sometimes with listening. Speaking was not too good and you can see that I still have some problems speaking. Reading I understand almost anything, it was 29."

Several students mentioned that the TOEFL iBT scores accurately reflected their language skills at the beginning of the year but that they had improved and would expect higher scores at the end of the year, as illustrated by S030, who thought that "I'm maybe better at reading and listening now. Before coming, yeah I think these are quite the same" [comparing the test scores to his skills].

Only eight EAP tutors/lecturers commented on TOEFL iBT's power to indicate language skills, the majority of whom regarded the test as a good indicator, while two of them expressed concerns about particular students' speaking and writing scores not reflecting their actual abilities.

TOEFL iBT as Indicator for Academic Performance and Success

The vast majority of students perceived TOEFL iBT as a good indicator of their academic progress and grades, particularly with regard to their written assignment grades and to their coping with the academic requirements to process input from lectures and texts. This view was supported by all tutors who commented on this aspect. Two students mentioned that their speaking scores were a good indicator for the grades they received for presentations, whereas three students commented on the speaking and writing scores not being very good indicators for how they coped with the academic speaking and

writing requirements. The data showed that students and tutors were aware of the complex relation between language competence and academic success, as the following excerpt from S005 (second interview point) illustrates:

> I think the TOEFL enables you to reach a certain standard. From then on you can improve if you have a high enough TOEFL. If you don't, probably you can't even achieve an average mark. . . . English is not the main driver . . . you need to know it but then you build on it. . . . You need to be clever enough to understand what you get from the content, out of all the information and concepts that you get, how to apply them to practical cases.

This stance was supported by several students and EAP tutors/lecturers, some of whom thought that English was a prerequisite but not the determining factor for academic success, because the latter was affected by many other variables.

Appropriateness of Entry Requirements

The vast majority of students knew about the minimum required TOEFL iBT test scores for entry to their specific courses, and no student stated that these would have been too low or too high. When asking the EAP tutors and lecturers about the appropriateness of the required test scores, 15 mentioned that the scores were appropriate, with several tutors recommending not to lower the entry requirements, because students needed to "hit the ground running" (T25), and the university's reputation was expressed, among other aspects, by the entry standards. Some tutors mentioned the importance of looking at the section scores, particularly writing and speaking, while others recommended making use of interviews in addition to the test scores. With regard to supporting students at the lower end of the required scores, the presessional manager stated that reports for these students would regularly be sent to their departments and that students would be encouraged to attend in-sessional classes. It is important to note that no EAP tutor or lecturer asked for the entrance test scores to be lowered.

Exploitation of Language Support

Next, we addressed Theme 2/RQ4 to analyze students' exploitation of language support and their attitudes toward improving English. Again, we also took into account tutors' perspectives on students' needs and exploitation. In addition, we drew on indications of students' struggling and not being well prepared, as reported earlier, to shed light on the question whether those students struggling in particular areas were actually seeking support. Table 30 gives an overview of the selected codes we analyzed here as well as the coding statistics. We used the codes to organize the findings.

Students' and Tutors' Perceptions of Presessional Courses

Of the 25 students interviewed, 10 had undertaken the presessional course, for 5 weeks rather than 10 weeks in all cases but one. While the majority of students who had not attended presessional offered no specific reasons why, two referred to lack of time, as they had been working until just before term started; one mentioned the additional fee involved as a deterrent; while another reported no perceived need to attend as his TOEFL iBT results had been good enough. Analysis of our database confirmed that all other interviewees who had not attended presessional had similarly met the TOEFL iBT requirements for admission.

Among students who had done a presessional course and offered reasons why, three explained that they had not quite met the TOEFL iBT requirements for admission and had thus been required either to follow the presessional route to entry or to retake TOEFL iBT, while for two visiting Japanese students, the presessional course was a mandatory part of the study abroad year. Others, on the other hand, had freely chosen to do the presessional course even though they had met the English-language requirement for admission to their degree studies. Reasons offered included a perceived need to improve English-language skills further and to adjust to living in the United Kingdom.

When invited to evaluate the usefulness of the presessional course, students varied in their perceptions, possibly reflecting differences in disciplinary language emphases. For students from the Centre for Applied Linguistics and the Department of Politics and International Studies, where assessment is largely based on extensive writing (of assignments and dissertations), the presessional course was perceived to be particularly useful for focusing on academic writing skills, conventions, and practices, such as how to use writing frames and discourse markers, how to reference and paraphrase sources appropriately, and how to avoid plagiarism. On the other hand, students from the Economics and Manufacturing

Table 30 Coding Scheme for Theme 2/RQ4

NVivo parent node	NVivo child nodes	No. of sources	No. of references
1. Presessional classes	1. Effectiveness	21	53
	2. Ineffective aspects	16	29
	3. Reasons for (non) attendance	29	43
	4. Tutors' perceptions	28	134
2. In-sessional classes	1. Effectiveness	8	21
	2. Ineffective aspects	3	7
	3. Reasons for (non) attendance	36	69
	4. Tutors' perceptions	20	64
3. University language support	1. Offers	61	299
	2. Exploitation	51	159
	3. Effectiveness	37	88
	4. Not using it	54	103
	5. Expectations	8	16
4. Active learning/improvement (by student)	1. Yes	58	242
	2. Awareness of need to improve	21	49
	3. No active learning	21	42
5. Facilitators and constraints		51	128
6. Making use of tutor feedback		24	54
7. English-language support outside university	1. Yes	3	4
(attending courses)	2. No	5	7
9. TOEFL test results and seeking support	1. Use(fulness) of test reports	62	364
	2. Test results as impulse to seek support	13	37
	3. Ideal report	47	119

Departments highlighted the effectiveness of the presessional course in helping with speaking practice and confidence, with adjustment to life in the United Kingdom, and with establishing friendships that then lasted through the year. It should be noted that the Economics and Manufacturing Departments have very large postgraduate cohorts, so the need for social confidence in speaking and establishing friendships may be more keenly felt than in smaller departments. For students interviewed two or three times during the year, these perceptions of what they found helpful about the presessional course remained remarkably stable. For example, one student commented that even for her dissertation writing, she found herself referring back to useful materials and phrases acquired from her presessional tutor.

At the same time, analysis of both student and tutor perceptions of the presessional course suggests that, on the whole, its potential for actually improving students' English-language abilities (as opposed to their academic skills or confidence) may be felt to be rather limited. This was a point explicitly acknowledged by one EAP tutor (T40), who has been teaching presessional for many years, and who also commented that Phase 2 was becoming too test focused (a view echoed by other EAP tutors interviewed). As one student observed, the period of presessional study (5 weeks for the majority) is too short for improving writing and speaking skills. This view was expressed too by an academic lecturer in manufacturing who highlighted the desirability for both phases (10 weeks) to be mandatory for students who fall short of the English entry requirement. In fact, as he noted, this had been the case in 2012–2013 but had resulted in a fall in student recruitment, leading the department to revert to making only one phase of presessional mandatory for such students, which in his view was not enough. A similar concern but somewhat different practice was reported by a lecturer in statistics, who said that his department preferred not to admit students conditionally via the presessional route. This was because the English-language "hurdle" (requirement of TOEFL iBT overall score of 92) was already perceived to be low enough, and the department could not feel confident that students who fell short of this hurdle would improve sufficiently by attending presessional.

Other students noted that presessional might not be that helpful for meeting the actual language demands (such as specialist vocabulary or genres of writing) of one's degree course and that much depended on the disciplinary expertise of the particular EAP tutor assigned to each presessional group as well as on the composition and language level of students in the group. Nevertheless, 4 of the 10 students who did presessional perceived it to have had a positive impact on their academic grades. For one of these students, this positive impact was particularly evidenced in the grade for his first assessed assignment, where he felt he had a competitive advantage over colleagues who had not done presessional.

On the whole, academic lecturers interviewed had few comments to make about the presessional courses. Many seemed unaware whether their students had done presessional courses (unless they heard about it informally from particular students), and most did not have sight of the presessional assessment reports for their own students or personal tutees. The exceptions were academic lecturers in the Centre for Applied Linguistics, who commented on the usefulness of the presessional reports for checking their personal tutees' English-language skills. For lecturers in other departments, there was a general perception that the business of looking at students' presessional reports, English-language proficiency, and entry scores was a matter for staff responsible for student admissions and not really their concern. As one lecturer (T25) in the Business School commented, having sight of information about students' English-language proficiency would make no difference because lecturers "make an assumption [about students] coming in that they've already been approved by the system as having the appropriate level of reading, writing, and understanding, etc.," and therefore would not expect to have to adapt their teaching practices to students' varying English levels. Academic staff interviewed in business and engineering also seemed to think that their departments had their own English-language screening procedures that did not involve the presessional route, though T58 (Engineering) further opined that many students he taught seemed to lack the necessary language and academic skills, which he attributed to the ineffectiveness of a screening process that relied on secondary information rather than, for example, face-to-face interviews.

This general lack of academic lecturer engagement with the presessional course and the assessment reports produced is recognized to some extent by the presessional EAP tutors interviewed. Among the six EAP tutors interviewed, opinions varied as to the usefulness of the assessment reports for target departments and academic lecturers, with doubts expressed as to how far they were actually read or acted upon. On the other hand, the EAP tutors generally agreed that the reports were genuinely useful for students in clarifying their individual strengths and weaknesses, boosting confidence, and (where relevant) identifying need for further in-sessional language support. The main aim of the summative test at the end of presessional was to evaluate whether "the student is capable and ready to undertake work at the university" (T45). The final presessional report contained a detailed section on strengths and areas to improve, along with test scores of the final tests. Where weaknesses were assessed, a letter was usually written to the department and in-sessional support recommended. Three students who attended presessional mentioned that they found the formative feedback very helpful and specific, while they perceived the test score results as somewhat less insightful.

Students' Exploitation of Ongoing Language Support

In each interview, students were asked if they availed themselves of any ongoing language support and to talk about their reasons and perceptions relating to such support. The support available within the university included in-sessional English classes and individual consultation offered by the Centre for Applied Linguistics, open to all international students; tailored in-sessional English classes provided by Centre for Applied Linguistics for specific departments; various kinds of seminars and workshops (e.g., academic writing and practices, research writing, professional communication, and dissemination skills) run by particular departments (e.g., Business School) for their own students or provided centrally by the university for all (home and international) postgraduate students (e.g., Academic Writing Program, Master's Skills Program, Research Students Skills Program); and informal conversation classes and language and cultural exchange sessions run by the Students' Union for home and international students (language buddy scheme, language café).

Of the 25 students interviewed, only three reported taking in-sessional language classes (confirmed also in our in-sessional attendance database). S008, who attended academic writing classes in Term 1 followed by dissertation writing classes later in the year, explained that she had been advised by a tutor on a predeparture course to attend in-sessional writing classes and acknowledged that she struggled with writing: "just the writing, my writing skill was not that good." This self-perception was reflected also in her TOEFL iBT, profile which showed writing to be her weakest score. However, while S008 found the academic and dissertation writing classes very useful, S014 (who also reported struggling with writing, even at the final interview, and whose writing score was similarly the lowest in her TOEFL iBT profile) felt that the academic writing in-sessional classes were too mixed in terms of students' English levels to be useful. For S039, the main focus was on developing pronunciation, listening, and interactional skills (reflected in low TOEFL iBT scores of 17 and 22 for listening and speaking, in comparison with scores of 27 and 29 for writing and reading). Aside from attending in-sessional pronunciation classes (perceived as very useful), S039 also reported taking advantage of the informal language buddy scheme provided by the Students' Union as well as other opportunities for social interaction (in his words, "partying" as a language practice strategy) to improve his conversation skills.

Among 10 students who gave reasons for not availing themselves of in-sessional language support, 5 did not perceive a need as they were managing well enough, although all talked elsewhere about some language-related challenges and struggles they experienced in relation to writing, speaking, or understanding local and international English accents. The other five students reported being either unaware of the available in-sessional support or not having time to make use of it. Three in particular wished they had been able to attend in-sessional classes as they felt themselves to be struggling with writing or speaking, even when interviewed toward the end of the year (e.g., S006: "I'm still feeling that it's a struggle sometimes when I want to write something and it wasn't coming out so well").

In relation to departmental or centrally provided support for academic and professional communication skills, eight students reported making use of this provision and, in all cases, commented positively on it. A point worth making here is that such departmental or central provision is not aimed at international students per se but at the whole student body, whether in a particular department or degree program (e.g., MBA) or across the university (e.g., Master's Skills and Research Students Skills Programs). This may make such integrated provision more attractive to international students (e.g., for the opportunities to share problems and experiences with home and international students, as S016 remarked) than in-sessional English-language support that may be perceived as largely remedial in function (i.e., in the words of S018, "if students really scored not that much good in TOEFL"). As a presessional EAP tutor (T45) observed, some students may feel they lose face if they seek out in-sessional support ("it's like a sign of weakness if they actually ask for extra help").

Interestingly, on the other hand, academic lecturers tended to be more critical and discriminating in their perceptions of what support international students needed. Although a few lecturers commented positively on departmental and centrally provided support for home and international students in relation to academic and professional communication skills, some perceived international students to have particular needs requiring more dedicated support. For example, T36 expressed rather negative views, commenting that the general introductions (to academic writing and academic practices) offered were "completely ineffective" in dealing with the language and cultural barriers presented by many international students in his department (Manufacturing). Although departments seemed to have systems for flagging up students with language issues needing support (e.g., through personal tutors or evaluation of a trial assignment early in the year), several academic lecturers voiced doubts as to how far international students made good use of the language support services available. For example, T38 (Statistics) observed that students do not seem keen to attend the dedicated language support classes offered, even when advised to do so, while T24 (Business School) commented that students may resort to such support only when they realize they need it, which may of course be rather late in the year. T58 (Engineering) acknowledged that students were often too busy with their main studies to have time to attend language classes. He then expressed misgivings about the initial screening process if some students fail parts of their course despite having met the English-language requirements because they have not had recourse to support: "So we put them in a situation, we put them in the lion's cage, without equipping them with any tools to fight the lions."

Students' Personal Strategies for Improving English

Although only a minority of students interviewed reported taking advantage of available language support services, all but two acknowledged that they actively sought to improve their English skills in various other ways. Indeed, some described language improvement as an explicit goal of their stay in the United Kingdom, while most identified immersion in an English-speaking environment as a significant attraction and facilitator for improving their language skills. Inevitably, there was considerable variation in the range of strategies students reported using to improve their English and variation in the level of specificity and metacognitive effort associated with these strategies. Some students spoke in rather general terms about everyday activities, such as watching television, reading, talking with friends, or exploiting opportunities for social interaction. On the other hand, several students elaborated more specific strategies they used to address particular weaknesses or develop particular skill areas. For example, S014 reported that she noted down new or interesting words she came across and also saved useful e-mail models for future adaptation. S012 explained that he studied how sentences were formulated in academic texts he read, paid attention to his own pronunciation when speaking and asked friends to correct it, and also memorized new words encountered when talking with friends and then made a point of reusing these words. In a similar vein, S016 described how he paid close attention to different ways of saying things in his interlocutors' speech, nicely demonstrating this strategy during the interview as he talked about it:

INTERVIEWER: Okay, so your main focus is on expression and vocabulary?

s016: Exactly. See now I didn't have the word "expression" in me. So I learnt from you now.

S004 reported that he revisited assignment drafts after a few days to correct and revise his writing before submission, having realized the importance of doing so upon receipt of a lower mark for an unproofread assignment. In this regard, several students also reported that they paid close attention to tutor feedback on assignments, with some noting that the feedback raised their awareness of certain language issues requiring attention. This was the case, for example, with S039, whose presessional tutor and academic lecturer (T56, T58) both corroborated his proactive behavior of paying attention to and acting on feedback.

Across the board, then, students appeared keen to improve their English skills, whether through need or desire, and most reported actively engaging in various strategies to this end. Many showed awareness of weaknesses or skill areas needing improvement, and some reported developing new kinds of metacognitive awareness either about language learning and use or about themselves as learners, through their experiences of engaging with English in a different environment. For example, S032 reported becoming more aware of the emotional and pragmatic context of English word use, while S004 noticed (through his part-time work as a steward in the university's Arts Centre) how focusing attentively on what interlocutors were saying made a significant difference in the quality of his understanding. With particular reference to perceptions of TOEFL iBT, S031 noted her growing awareness that she needed to master a more flexible range of structures in her academic writing than the "fixed forms of writing" typically practiced by Chinese students for TOEFL iBT.

Among perceived constraints affecting students' strategic efforts to improve their English, lack of time was commonly cited, as well as fewer than expected opportunities to interact with British students as opposed to other international students and a tendency to socialize with students from one's own country. This tendency for students to stick within their own cultural groups was highlighted in a somewhat critical vein by several academic lecturers, particularly in the Business School and Manufacturing Department, which have large cohorts of Chinese students. However, as T36 (Manufacturing) acknowledged, the fault lay as much with the British students for not integrating, and perhaps with the lack of a departmental "buddy" system to pair up British and international students in a more supportive fashion.

Use and Usefulness of TOEFL iBT Test Reports

When we asked students what use they made of the TOEFL iBT test reports, how useful they found the reports, and what an ideal report would look like, a trend emerged for students to mainly look at the reports to see whether they had reached the required minimum scores. In this case, the majority of students stated that they would not do anything else with the reports. Those who did not reach the required threshold used the test reports to see what they needed to improve for a second attempt to take the test. When this attempt yielded high enough scores, no further notice was given to the reports. Interestingly, most students showed a good level of awareness of their weaknesses, but only one student (S037) reported having actually made use of the test reports to seek targeted support for the weakest area of speaking (24):

INTERVIEWER: Did your TOEFL test result influence you to decide to come and attend the presessional course or seek language support?

s037: Well yeah, because I wasn't satisfied with my result from the speaking section. And in presessional they do have the preparation for presentations, so that's all the reasons why I decided to come to presessional, to improve those weak area of my English language.

One possible reason for not making wider use of the test reports could lie in the fact that the vast majority of students found the reports too generic. Students stated that they would prefer individualized, personalized feedback on weaknesses and on what and how to improve, particularly for writing and speaking. Yet many students were aware that such individualization may not be feasible in a large-scale testing context. S030 illustrates the perception of the majority of our students:

s30: Because that is why I didn't read it, I always think okay they have copy paste. . . . It's nothing helpful, but if they can give I don't know, especially in the speaking and writing part, this parts were more difficult than listening and reading. In this part I used that it's better to have a feedback that is, that comments on our mistakes. Rather than saying what we made right, what we made wrong.INTERVIEWER: So if it could be more individualized, you think that would be more useful?

s30: Yeah, and I am not sure they can do that.

This awareness of what is feasible in a university entrance test is underlined by S034:

> Well the most useful test report would be of course a personalized one, but since there's millions of people, or like thousands of people taking this test, it's impossible to give a personalized one to everybody. I think they've done a good job with this one.

In line with S034, many students stated that the TOEFL iBT test report was useful for the purpose of the test, that is, allowing them access to university.

The perception that the TOEFL iBT test reports offered rather generalized feedback only was shared also by the EAP tutors and academic lecturers interviewed, the majority of whom had not come across these reports before. Several did acknowledge the potential usefulness of the descriptors in interpreting TOEFL iBT scores in a generic sense, especially for staff new to admissions and recruitment roles. In particular, T63 (Business School) commented that staff involved in application screening and recruitment interviews for the MBA program "should sort of have a copy of this [report]" because "it's really crucial that they understand what it is they're looking for." Generally speaking, it appeared that staff involved in screening applications concerned themselves only with the test scores without reference to the reports, as acknowledged, for example, by T13 (an admissions tutor in the Manufacturing Department).

In terms of the reports' potential usefulness for students, some lecturers wondered aloud how far students would actually read and engage with the advice provided. Several commented that the generalized nature of the feedback would not help students in analyzing their individual language performance, such as identifying specific types of grammar mistakes they are prone to making, as opposed to simply recognizing that they make a lot of mistakes (T43). As a consequence, as T30 commented, it would be difficult for students to act upon the advice given, while others noted the desirability of more specific reference sources or links to resources (e.g., videos) for addressing particular skills or subskills. In relation to subskills, one lecturer (T48, Politics and International Studies) noted the need to include a focus on listening and interactional skills in large-group settings and not only one-to-one interactional skills. Another lecturer (T31) in the same department commented that students may benefit from more discipline-specific language advice, such as recommendations to engage with key readings in one's subject area, rather than reading a variety of academic texts in general. In her view, such discipline-specific feedback would be especially relevant for postgraduate international students, though perhaps less important for undergraduates.

Role of Language for Academic Success

To analyze students' and their tutors' perceptions of the role of language for academic success, we drew on Coding Theme 3 to answer RQ5: What role does language play in academic success? Table 31 gives an overview of the codes and the coding statistics we analyzed here.

Students' Perceptions of the Role of Language in Academic Success

We asked students what role the English language played in their academic life, assignments, and marked presentations. Here differences between departments became apparent. Students from the Engineering, Manufacturing, Statistics, and Mathematics Departments, as well as the Department of Economics and the Business School, reported that content and the ability to express one's opinion counted more than the style or correctness of one's English, as long as the message was getting across: "I think here they focus more on the content. But obviously it helps a lot if you, if you write well" (S017). Students acknowledged that being proficient in English helped them to make academic progress, but the grades ultimately depended more on knowledge, content, general writing, or presentation skills; marks would only be deducted for incomprehensible passages: "They do not penalize us for grammar mistakes as long as it's comprehensible" (S004). The reports by students in the Centre for Applied Linguistics and the Departments of Sociology and Politics suggest a somewhat different situation: Here language seems to play a much more important role as, for example, S038 expressed: "It's not hard to write the actual essay but it's just the grading is quite strict so if your English is really bad they might not. .. they might be kind of strict on that, I'm not sure."

Table 31 Coding Scheme for Theme 3/RQ5

NVivo parent node	NVivo child nodes	No. of sources	No. of references
1. Students' perception of role of language in assignments/exams/drafts, etc.	n/a	38	100
2. Tutors' attitude toward and emphasis of language (teaching, supervision, etc.)	1. Student's perception of tutors' attitude toward language	21	44
	2. Tutor's attitude toward/perception of language	39	259
3. Feedback from tutors on assignments, drafts, etc.	1. Focus	85	322
	2. Quality, usefulness	24	55
	3. Issues with feedback	21	37
	4. No feedback	24	53
4. Assessment criterion for language	1. Awareness, attitudes	53	196
	2. Not aware of it	7	13
5. Assessment practices		58	182
6. Student's academic progress	1. Effect of English proficiency on progress	62	155
	2. Academic grades	55	127

Across all departments, several students acknowledged the intertwinedness of content knowledge and linguistic expression and found it "difficult to separate the two" (S019). One student mentioned that it was during presessional that the focus was on language, while his academic course focused on content. Interestingly, students' perceptions of the role of language did not change over the course of the year, but some students perceived that lecturers seemed to have been more lenient at the beginning of the year.

While students generally perceived that lecturers showed leniency toward international students' imperfect English, they acknowledged that the role of language for their academic success and grades ultimately depended on their lecturers' attitudes: "It depends on the lecturer and the tutor, because some of them are aware that we are foreign and they are quite relaxed with the English; others are more strict" (S039). This is also expressed in the feedback students received on their linguistic performance: It differed widely, not only from department to department but also from lecturer to lecturer. Students also were well aware of the fact that all faculties have a marking criterion to assess "presentation," which includes the quality of language, yet lecturers seemed to weight this criterion quite differently. Next, we explore the attitudes toward language as reported by EAP tutors and lecturers first, before we turn to feedback and assessment practices.

Tutors' Attitudes Toward Language

EAP tutors reported, as was to be expected, that their main focus was on academic language, including critical reading skills, summarizing different sources for a writing project, listening to academic lectures, giving presentations, working on intelligible pronunciation, punctuation, academic style and register, paraphrasing, avoiding plagiarism, and developing students' voice as authors. For in-sessional classes, EAP tutors reported giving formative feedback only on the aspects covered in class, with no assessment taking place. With regard to presessional courses, all six EAP tutors reported giving detailed formative and summative feedback on all aspects of language, both in written and oral form; in addition, they offered individual and group tutorials.

We then analyzed lecturers' attitudes in Applied Linguistics, Sociology, and Politics, the departments where students had reported that more focus was given to language. Not surprisingly, perhaps, the seven native speakers and two nonnative speakers expressed very diverse views. In line with students' perceptions, the nonnative lecturers tended to make more allowances and to adjust their teaching and communication styles, but there were also native speakers who very consciously prepared handouts and accommodated for students' linguistic needs. While lecturers seemed to agree that teaching and supervision should focus on content, not language—as T42 put it, "there is probably only so much that can and should be required for us in terms of providing the English proficiency and support"—some lecturers reported that they would point out linguistic issues in one-to-one situations. Several native speakers commented that they would not make allowances in their teaching, supervision, and communication styles because standards needed to be kept up and students were working for a degree from a U.K. university. This mirrors students' perception in the selected departments that language was given some importance. Several lecturers pointed out that they usually did not have to make linguistic

allowances because students in general came in at the right language level and coped well with the language requirements. T32, for example, "never felt any major difficulties with teaching that would derive from language."

As indicated in the student interviews, lecturers in the Engineering, Manufacturing, Economics, Statistics, and Mathematics Departments and in the Business School indeed showed the tendency to give language a somewhat less prominent role, focusing more on content and application of theories; T21, for example, was "looking for sophistication of thought and analysis rather than dotting the *i*'s and crossing the *t*'s or knowing a particular idiom." Similar to the lecturers in the more language-focused departments, most lecturers showed a high level of awareness of how to adjust their teaching and supervision to the needs of international students, mainly by adjusting speed, accents, and simplicity of English, but many reported that large and heterogeneous student groups prevented them from adjusting their teaching style to individuals' needs. Comparing native speakers to lecturers for whom English is a foreign language (15 vs. 6), we did not find emerging differences: In both groups, there were lecturers arguing for keeping up standards and not making any allowances as well as those who acknowledged and allowed for the additional challenges international students face. T24, for instance, intended "not to penalize their English because I think it's hard enough all the other things they're coping with really." A few lecturers commented on the importance of having a good level of English also for technical subjects, such as T34, who felt that "with postgraduates there's a sense in which they should be producing things that are of a high standard, especially PhD students, and if they can't write good academic English, then how are they going to progress, in the future?" while several lecturers stated that it was the technical skills that mattered most. As with the first lecturer group, the lecturers in the more technical departments did perceive that students were admitted at the right level of English and that it was only a very small minority who struggled with the linguistic requirements. T37, for instance, "never really had to be concerned about the language capability of the students" and was "happy with the [admissions] system."

The Role of Language in Lecturers' Feedback on Academic Work

Students across all departments reported a tendency for lecturer feedback to focus on content and on expected structures for written assignments; hardly any student reported to have received feedback on oral presentations, and there was no instance in our data of feedback on tests and exams. About half the participants across all departments reported that lecturers did not give feedback on linguistic issues, while the other half reported to have received such feedback. From students' reports, it appeared that they received feedback on linguistic issues when and where needed; if the language was good, no feedback on language was given. As S036 put it, "you don't get feedback for your English, you get the feedback for your piece of work. If you have got some English problems or maybe because the assessor can't read, it could happen." This trend was reported across the year, with only two students mentioning that they received linguistic feedback only at the beginning of the year. The majority of students perceived the feedback they received as helpful and as facilitating their academic improvement. Students characterized the following aspects as particularly helpful: specific rather than generic feedback on selected language aspects; feedback that clarifies expectations, conventions, and structure; and feedback on how to develop a convincing argument.

Notwithstanding the fact that student reports did not reveal differences in feedback behavior for different departments, we looked separately at the two departmental groupings used earlier to ensure that we captured every possible angle. Lecturers in the first group (Centre for Applied Linguistics, Departments of Sociology and Politics) reported a variety of feedback practices, in line with their previously outlined attitudes toward language. Those who thought that language did not play a major role usually focused their feedback on content, only pointing out linguistic issues if the meaning was incomprehensible or language use was inappropriate. One lecturer would comment on "careless mistakes" (T43); another, however, would instill confidence in students' language use by giving "encouraging feedback" (T48). Those lecturers for whom language was important would generally give feedback on linguistic issues, in parts even going so far as to "correct the English language" (T28). It is noteworthy that two lecturers (T42, T44) mentioned that their feedback focus shifted to content only during dissertation supervision. We could not find differing trends across native (L1) and nonnative (L2) speakers.

The second lecturer group in the more technical, quantitative departments showed a slight tendency to focus in their feedback more on content and projects and not so much on language issues. T25 summed this up succinctly:

> Tutor feedback on linguistic issues: If necessary, if asked for, and if it's seen as particularly problematic, it will probably be raised, but it's not something, to my knowledge, which is a requirement. It's more about the quality of the

> work, their understanding of the subject matter, and not the linguistic issues; but if linguistic issues are the thing that's interfering with their communication, then it probably will be raised.

Nevertheless, five lecturers (two nonnative speakers) stated that they would give explicit feedback on linguistic issues, including corrections, if needed. Two lecturers remarked that they would give feedback on language in a "discreet way" (T35) or on a "one-to-one basis" (T26). Several other lecturers mentioned that they would only point out severe linguistic issues and direct students to experts for language support. As with the first group of lecturers, we did not find differences between L1 and L2 lecturers in their reported feedback behavior.

The Role of Language in Marking and Assessment of Academic Assignments

The vast majority of students across all departments in our study were aware of an assessment criterion relating to presentation and language. We looked separately at the two departmental groupings that emerged earlier. Interestingly, the seven students from Applied Linguistics, Sociology, and Politics answering this question showed a broader range of perceptions toward the question whether language is assessed than did the 15 students from the more technically oriented departments. Here there was a clear trend mirroring what students had reported with regard to the importance of language in general: Language would only play a role in assessment if it was so poor that it impeded comprehensibility; otherwise, lecturers would primarily assess the academic content and quality of students' work. Nevertheless, several students acknowledged the appropriateness of giving language a certain weight (factors between 5% and 10% were mentioned), not least "because it's an English university. We have to be able to communicate our results and our explanations in the best English that we can" (S002). Several students mentioned that the assessment was anonymous, hence lecturers could not make allowances for nonnative speakers. Interestingly, toward the middle and end of the year, several students noted how important it was to take enough time to proofread and check the language before submitting an assignment; otherwise, it would affect the grade, as S018 reported: "If I am writing like in a hurry I generally tend to do some grammatical mistakes.. .. The feedback given back by my assessors is for grammatical mistakes in that they deduct marks from my assignment."

As was to be expected, lecturers' attitudes were also reflected in their marking and assessment procedures. Interestingly, lecturers from different departments reported that there was no policy within the department or the faculty on how to interpret the marking criterion "presentation," which, among other aspects, focuses on quality of language. Ultimately, it seems, each lecturer interpreted, applied, and weighed this criterion as he or she saw fit and appropriate for the module to be assessed.

Lecturers in the group of departments with a social sciences focus (Applied Linguistics, Sociology, Politics) showed a huge variety of approaches to marking, ranging from being very lenient in cases where English was not the students' first language to showing a very strict attitude toward correct language use, placing importance on maintaining standards. Several lecturers mentioned that language did have an impact on marking, positive as well as negative, but that it was difficult to disentangle content and form, as they influenced each other. Quite a number of both native and nonnative lecturers stated that language was not the decisive criterion, with one nonnative lecturer (T32) going so far as to state, "I don't think English ability should be a criteria for assessment, because that would introduce. .. some problems. .. if I made allowances for the fact that people are not native speakers. I treat them all equally in the assessment." The latter remark mirrored the perception of some students, while other students and lecturers perceived that international students' English received appreciation, as T48 described it:

> I've never heard any sort of explicit discussion about how we take international students' language ability into account. So I can only speak personally. Which is, it's very clear to me if English isn't someone's first language. And I will, I will read the essay differently, in the sense that I will be trying to really, making the effort to grasp what that student is getting to, and appreciating that it's not their first language. So I don't personally this doesn't have a big impact on their grades.

The variety of approaches reported by the lecturers for the first group of departments does reflect the range of students' perceptions as outlined earlier. Nevertheless, lecturers seemed to agree that the criterion of presentation or language carries less weight than the criteria focusing on content, analysis, and critique.

Lecturers in the second group of departments (Manufacturing, Engineering, etc.) showed a clearer tendency toward arguing that language did not have an impact on assessment and grades as long as students managed to communicate their understanding and to clearly and comprehensibly make their point. Generally, lecturers stated that it was the content and application of theories that mattered and that they would be tolerant toward linguistic issues in their marking. Marks would only be deducted if the work was not understandable, if students could not demonstrate their conceptual understanding, or if they failed to develop coherent arguments. Many lecturers, however, acknowledged the interrelatedness of content and language, as illustrated by T34:

> It can be difficult marking written work because you don't know if technically they have understood what they're writing or whether it's just their English is bad? So there is just a problem that you want to only mark them on technical content, but if they can't communicate that, then it's very difficult.

Another lecturer (T13) acknowledged the possibility that the quality of English may indeed affect the grade, albeit unconsciously, while yet another lecturer (T61) stated that even native speakers sometimes did not manage to convey their ideas in a comprehensible way and would be marked down for it. With regard to differences between L1 and L2 lecturers, our data did not show any differences in their marking approaches. Interestingly, one lecturer mentioned the tendency to be more lenient at the beginning of the year, underpinning what some students had noted earlier. In brief, T62 summed up the role of language succinctly: "What role the language plays is, the better you express the ideas the better the mark."

Effect of Students' English Proficiency on Their Academic Progress and Grades

When we asked students whether their English proficiency had an effect on the academic grades they received, their responses varied to a certain degree, with about half the students not perceiving a large effect and the other half thinking that English did have an effect on their grades. It was difficult to compare students from the more language-oriented departments with those from the more technically focused ones, because the majority of students answering this question came from the latter departments.

According to our data, the vast majority of students recognized that a higher English proficiency will most likely have a positive effect on academic performance, as S006 put it: "If someone will have a better proficiency they probably can explain the same thing that I'm explaining in a better way. .. in a way that the reader can understand easier. So that might affect the grade in some way." One student mentioned a threshold of English beyond which the language did not influence academic performance. In addition, many students acknowledged that content knowledge and knowing about the expected structure of academic work had an impact probably larger than language, as S005 illustrates: "Whereas I might also have been less proficient in the proper academic subjects whilst still having a good English I would still have got lower marks anyway."

The majority of our students perceived that they had to make more effort than a native speaker, spending more time on reading sources and on proofreading. If they spent this additional time on proofreading and editing their written assignments, they felt that their language proficiency did not have a negative effect on their academic grades, as illustrated by the following quote by S004:

> Usually I write an assessment or a paper, and then I wait a few days and look at it again and I have to rewrite some pieces that don't make quite as much sense in English as I thought they did. I know that I didn't do that for one paper and I got a lower grade than I used to, so I can see that really playing a role.

It was only for oral presentations that two students mentioned they received lower grades, attributing the grades to being nervous and under pressure, which would probably not have been the case had they been able to present in their L1.

Furthermore, many students recognized their improvement over time, both for the expected structure and their writing skills, as S034 mentioned in the second interview: "I think it is just the way of doing your research and getting used to the way they want things done, but also my writing skills are developed every day, it gets better every day." It appeared that by the time students were working on their dissertations, they were well acquainted with the academic expectations and had developed their writing skills accordingly.

With regard to the lecturers' perceptions, again we have an imbalance between the two departmental groups. The four lecturers from the more language-oriented departments stated that there was a threshold of proficiency below which work was not comprehensible, which would of course affect the grades. One lecturer (T32), however, stated that "no student so far of mine has had a major problem or had major difficulties with language. .. in a sense that language was a decisive factor in their performance," whereas another lecturer (T42) mentioned that to receive the highest grades, the work "has to be of publishable quality, meaning if it's written in not perfect English, it's clearly not publishable quality, so you can't go up there."

The lecturers from the more technically oriented departments seemed to share this diversity. Although some lecturers acknowledged the earlier mentioned threshold of proficiency necessary to get good marks, as, for example, T38 put it — "if the sentences don't make any sense, you can't write in any kind of precise way, and in that case you just can't obtain a good mark" — other lecturers placed much more weight on content and understanding, such as T36, who did not "think the language is such a barrier.. .. It doesn't have to be perfect English. But if they understand or not, that is going to have an effect."

In sum, students and lecturers across departments seemed to agree that good academic marks require an ability to express one's understanding and to structure a convincing argument or case, which in turn requires a certain level of English proficiency and academic writing skill. Although nonnative students may have to put more time and effort into their academic writing, several participants mentioned that being a native speaker does not guarantee high academic grades.

Interpretation and Discussion of Findings

We now discuss our findings with a view to answering each research question in turn, drawing on quantitative and qualitative findings from both strands.

RQ1: Relation Between TOEFL iBT Scores and Academic Success

We first address RQ1, the relation between the language skills reported by the TOEFL iBT scores and students' subsequent academic performance as expressed in final academic grades. With regard to the question whether different subgroups of students show differing profiles in their TOEFL iBT section scores, like Bridgeman et al. (2015), we found that Chinese students exhibit a slightly different profile of TOEFL iBT section scores from the rest of the population, yet it was much less pronounced than what Bridgeman et al. reported. Hence we found no grounds upon which to recommend paying specific attention to certain groups for conspicuous profiles. Whether this trend is representative for all students coming to the United Kingdom with a TOEFL iBT as a language entrance exam for academic studies would need to be investigated further.

In line with findings reported in Cho and Bridgeman (2012) and Bridgeman et al. (2015), we found small but significant correlations for all TOEFL iBT scores for the total sample. Following Bridgeman et al.'s approach, we found that "peeling the onion" and looking into different subgroups for correlation patterns yielded a more nuanced picture. Examining correlations by faculty suggests that the strongest relations between TOEFL iBT scores and academic outcomes are found in the arts faculty (significant for TOEFL iBT writing, overall, and speaking), followed by science (weaker but significant for all but the writing scores), and finally social science (weak but significant for TOEFL iBT speaking, listening, and overall). Grouping students by selected departments with a quantitative focus, on one hand (selQUANT), and a social science focus, on the other (selSOC), as the interview data suggested, we found that only selQUANT students showed small correlation coefficients for speaking, listening, and overall TOEFL iBT scores, while there were no significant correlations for selSOC departments. These results could cautiously be explained in light of the interview findings, which suggested that language plays a greater role in social sciences, where students reported they received feedback on linguistic issues and their academic lecturers paid particular attention to language. Hence initial differences in social sciences students' TOEFL iBT scores may be leveled out by the focus on language from both tutors and students, so that TOEFL iBT scores and final academic grades show no significant relation for selSOC students. However, it has to be noted that the selSOC group contains very few students. In the much larger social science group, the TOEFL iBT correlation is substantial (.41 adjusted ρ). Unfortunately, we did not have a single student from the arts faculty in our interview sample so that we cannot shed more light on the findings for this faculty.

Grouping students by nationality also yielded interesting results: The Chinese subgroup showed the strongest correlations, significant for speaking and overall scores, while for the Indian subgroup, the correlations were only significant for listening scores, broadly in line with the findings reported by Bridgeman et al. (2015). In addition to these two groups, we also examined the subgroup of German students, who showed the smallest correlations, none of which was significant. For our three subgroups, it seems that the group of students furthest apart from the English-language and U.K. academic culture shows the strongest relation between language and academic success, while for students from a Western educational background, English-language proficiency may account less for academic success. However, we have to concede that sample sizes may have an effect on the correlation findings.

The question of what effect additional language support has on the final academic grade can be answered as follows for our sample: While students who attended pre- or in-sessional classes do have significantly lower TOEFL iBT scores than students without support, as indicated by a *t* test significant at the .001 level, a chi-square test (*p* value at .136) did not indicate that there are differences in the academic outcomes between the two groups. The fact that we could not detect statistically significant differences in the final academic grades between students with language support and those without implies that the additional language support is effective in supporting students at the lower end of the TOEFL iBT score range, with the result that they do not seem disadvantaged with regard to academic outcome in comparison to their fellow students coming in with higher TOEFL iBT scores.

RQ2: Predictive Power of TOEFL iBT Scores on Academic Success

We now turn to the discussion of our findings with regard to RQ2, the potential of the TOEFL iBT scores to predict students' final academic grades and the effects of selected variables (academic disciplines, nationality, and additional language support) on the predictive relation between TOEFL iBT scores and academic outcome. The expectancy graphs we employed following Cho and Bridgeman (2012) give a somewhat clearer picture of the relation between certain TOEFL iBT score bands and certain academic grades than the correlations reported previously, showing a trend for students in higher TOEFL iBT bands to achieve higher academic grades, while the few students who failed or received a lower academic grade were most likely to be found in the bottom TOEFL iBT score band, in line with Cho and Bridgeman's findings. It seems that language plays a remarkably determining role in the arts disciplines (as indicated by the correlations), while this trend is less pronounced in science disciplines and even less so in social sciences. The faculty-related findings are difficult to compare to Cho and Bridgeman's findings, because they grouped students according to a different scheme. With regard to our findings for the social sciences disciplines, this result may be somewhat contradictory to intuition, but as indicated earlier when discussing the correlation results, interview findings suggest that the particular attention paid to language in some social sciences disciplines may account for the trend that initial differences in language proficiency are leveled out and hence do not have an effect on academic outcome.

OLR models revealed the strongest predictive power for the TOEFL iBT overall scores and nationality, whereas taking TOEFL iBT section scores, faculties, or additional language support into the model did not add much predictive power to the simplest model of using TOEFL iBT overall scores only. In comparison to the findings of Van Nelson et al. (2004), we found a stronger predictive power of TOEFL iBT on final academic grade. In our case, it seems that nationality as a proxy for students' first language has the strongest predictive power, as was already implied by the correlational analyses. While this result could be regarded as an indication toward setting different entrance requirements for different nationalities, taking into account the nearness or distance of languages and academic cultures to the English language and the U.K. academic culture, this would be a problematic recommendation insofar as it could be regarded as discriminating between students based on their nationality. Hence we would not recommend differentiating between students' nationalities when setting entrance score requirements. Rather, we would recommend offering additional support for students coming from languages and academic cultures with more distance from the United Kingdom, particularly in the beginning of their studies, to provide them with equal opportunities. We discuss implications for additional language support in the RQ4: Exploitation of Language Support subsection.

The findings from Strand 2 generally support the predictive power of TOEFL iBT scores found in Strand 1. The vast majority of students in the interviews perceived TOEFL iBT as a good indicator of their academic progress and grades, particularly with regard to their written assignment grades, and to their coping with the academic requirements to process input from lectures and texts. This view was supported by all tutors who commented on this aspect and by the tutors answering the questionnaire, where TOEFL iBT was rated as a fairly good predictor of academic success.

RQ3: Linguistic Preparedness

We next discuss our findings from Strand 2, integrating these with the quantitative insights gained from Strand 1. RQ3 focused on the role TOEFL iBT plays in students' and tutors' perceptions of students' linguistic preparedness for academic studies—specifically, their perceptions of whether students are prepared for the linguistic demands of their academic studies and whether TOEFL iBT usefully contributes in this regard.

Generally speaking, students in our sample reported being well prepared by their prior English-language education to cope with the linguistic demands of their academic studies as well as everyday life in the United Kingdom. At the broad level of linguistic preparedness, this finding from self-report data in Strand 2 is thus coherent with the linear regression analysis in Strand 1 showing that even students admitted with low TOEFL iBT scores coped successfully with their academic courses and in many cases achieved good grades. Within the Strand 2 interview data set, students' perceptions that they were coping well were moreover confirmed by the final academic grades achieved (pass, merit, distinction) by all those following 1-year taught master's programs (as opposed to PhD programs).

Interestingly, the questionnaire data suggested that students' perceived levels of coping tended to increase as the year progressed and as they settled into their academic and living environments. The interview data shed light on students' language-related perceptions during the early period of adjustment, where frequent reference was made to initial challenges in attuning to different speaking accents (British as opposed to American, British regional, various international), engaging with lexically and linguistically complex academic texts, participating effectively in group work, and producing writing of the necessary standard for academic course work. For two thirds of the students interviewed, the process of linguistic adjustment extended also to everyday transactional and interpersonal communication skills, such as those associated with opening bank accounts or interacting with native speakers of English whose speech was fast paced, strongly accented, or colloquial. In relation to most of these language skill areas, initial challenges and struggles were generally perceived to have subsided by the second term as students gained experience and confidence in using English for their academic studies and everyday life. However, writing skills remained an area of concern for most students throughout the year, even for those following courses of study (e.g., economics, manufacturing) where assessment was not strongly writing based.

Broadly speaking, students' positive perceptions of their linguistic preparedness were consistent with the views expressed by academic lecturers, though a caveat here is that where lecturers were not commenting on a particular student of theirs in our sample, their perceptions related to international students in general rather than to the subset of students admitted on the basis of TOEFL iBT scores. Apart from one lecturer in the Department of Engineering who felt that many international students lacked the necessary language skills for courses he taught, there was a general perception among teaching staff that students were linguistically sufficiently prepared for their academic studies. Areas where students were perceived to struggle initially related to more extended forms of academic language production and reception, such as giving presentations, writing assignments, and listening to lectures, rather than to general interactional skills. These observations are supported by the research literature, which similarly points to writing assignments and giving oral presentations as representing particular challenges for international students (e.g., Woodrow, 2006; Zappa-Holman, 2007). In terms of giving presentations and listening to lectures, students' ability to engage fully in the large-group academic setting was felt to improve as they gained confidence and experience. In relation to writing skills, while lecturers' views varied across and within subject disciplines, there was a common perception that many students had difficulty producing well-written work in the early part of the year but that most did improve and learn to write in a satisfactory manner appropriate to the academic discipline and genre.

In short, across both student and tutor self-report data on linguistic preparedness for academic study, writing skills emerged as an important concern that remained an issue for some even toward the end. Interestingly, writing was also an area where the contribution of TOEFL iBT to students' linguistic preparedness was perceived in rather mixed terms, suggesting some complexity in the perceived relationship between TOEFL iBT performance and academic performance. On the positive side, students generally felt their TOEFL iBT writing section scores to be reasonable indicators of their language skills, though not as effective indicators as their listening and reading section scores (but more so than their speaking section scores, where the computer-based mode of test delivery and associated time pressures were felt to affect speaking performance). Moreover, nearly all students interviewed regarded TOEFL iBT as a good indicator of their academic progress, particularly as reflected in their grades for written assignments and in their associated ability to process academic input from lectures and texts. In this regard, students' own perceptions are consistent

with the findings from Strand 1, pointing to a general positive association between TOEFL iBT scores and academic outcomes.

However, in terms of the actual writing skill set needed for their academic studies, students' perceptions of how well TOEFL iBT contributed to their linguistic preparedness were less positive. TOEFL iBT writing tasks were perceived to be rather different in style, length, and convention from the kinds of writing tasks students faced in their postgraduate degree courses, and thus their value as preparation or training for academic study was felt to be limited. On the other hand, when invited to reflect back at the end of the year, few students remarked that they would change their approach to preparing for TOEFL iBT or for their academic studies if given the chance again. This suggests that, from the perspective of linguistic preparedness, the perceived discrepancies between TOEFL iBT writing tasks and academic course work writing tasks are not large enough to raise serious concerns for students, particularly when TOEFL iBT scores in themselves (including writing section scores) are regarded as effective indicators of both language skills and academic performance. Because all master's students in our interview sample did indeed go on to complete their studies successfully (most with merit or distinction), it would seem that they were able to bridge the gap between TOEFL iBT writing tasks and academic course work writing tasks without too much difficulty.

One possible interpretation here is that the process of bridging this gap is not so much a linguistic issue (i.e., improving one's English-language competence) as it is a matter of acculturation into the academic writing practices and conventions relevant to a particular subject discipline and genre. As one student (S005) astutely commented, "English is not the main driver" for academic progress once the necessary language threshold (defined by TOEFL iBT entry requirements) is achieved. At the university in question, it would seem that the language thresholds are set at appropriate levels for postgraduate studies in different disciplines, as reflected in the Strand 1 findings and also in students' and tutors' perceptions in Strand 2. Once students meet this language threshold, development of academic skills and support in academic acculturation may be more important (than linguistic improvement) for academic success (cf. Floyd, 2015). It is in light of this observation that we now turn to discussing students' exploitation of language support.

RQ4: Exploitation of Language Support

RQ4 concerns students' exploitation of language support and whether seeking support is associated with linguistic weaknesses perceived by students or tutors or reflected in TOEFL iBT section scores. Strand 1 findings revealed that students who had additional language support came in with significantly lower TOEFL iBT scores, yet they did not show significant differences in their final academic grades.

Strand 2 findings similarly pointed to TOEFL iBT section scores as an important indicator of decisions to take up language support, either where students had not met the TOEFL iBT requirements and opted for the presessional route to degree course entry (instead of resitting TOEFL iBT) or where students attended in-sessional classes to address particular skill weaknesses (as reflected in their TOEFL iBT section scores). However, although nearly all students interviewed acknowledged language-related struggles and weaknesses, only a small proportion actually sought in-sessional language support. While lack of time and lack of awareness (of in-sessional classes) were reported as reasons for not seeking language support, it was notable that there was proportionately higher take-up of departmental and centrally provided support for academic and professional communication skills. This departmental and central provision was aimed at all students (in a particular department, degree program, or set of degree programs) rather than targeted specifically at international students from non-English backgrounds. As suggested by our data, the "inclusive and non-stigmatizing" (Klinger & Murray, 2012, p. 37) nature of this curriculum-based provision may make it more appealing to international students, because it does not carry the association of remediation (of language deficits) implicit in seeking in-sessional support. Moreover, in light of our discussion of RQ3, it may also be the case that international postgraduate students recognize the necessity of acquiring relevant academic and communication skills for their degree courses and associated professional domains — that is, a set of academic literacies going beyond the language threshold of TOEFL iBT entry scores.

This interpretation is also borne out to some extent by students' and EAP tutors' perceptions of presessional courses, which tended to highlight beneficial factors other than linguistic improvement per se, such as relevant grounding in academic writing conventions and practices (e.g., how to cite and paraphrase sources and avoid plagiarism), acculturation into the academic and living environment, or boosting social confidence and establishing supportive friendships. Despite the fact that presessional courses are offered as an alternative entry pathway for students who do not meet the English-language requirements for their degree courses, it remains questionable whether they can be effective in raising students'

language competence, as opposed to facilitating their academic adjustment and progress in other important ways. As reported by Floyd (2015) in her research comparing international students who met the English-language requirements for university entry in Australia and those following an EAP entry pathway, it may be the case that learning in academic skills (through a presessional course) may help to equalize the academic performance of these students with that of their more linguistically competent peers. Our Strand 1 findings certainly suggest that students entering with lower TOEFL iBT scores and taking presessional courses were not disadvantaged in their final academic outcomes in comparison with those entering with higher scores.

In terms of whether students actually improved their English-language skills through the year, our data do not really allow us to make any inferences. Among students and tutors, there was a general perception (or assumption) that language skills did improve through the year, as reflected in the interview data and second student questionnaire data. For some students, improving their English was an explicit goal in coming to study in the United Kingdom, while nearly all reported using various cognitive, metacognitive, and social strategies to develop and practice their skills. However, some concerns did emerge around limited opportunities for interaction and integration with British students, which is recognized as a widespread issue across the internationalized HE sector (e.g., Education Intelligence, 2014).

RQ5: Role of Language for Academic Success

When examining the role of English language in academic success, differences between disciplines became apparent, as was to be expected from the literature (e.g., Bridgeman et al., 2015; Cho & Bridgeman, 2012) and from the findings in Strand 1. Students from selected departments with a quantitative focus (Business School, Mathematics Institute, Manufacturing Group, Economics, Statistics) reported that content and the ability to express one's opinion played a greater role for progress and high grades than the style or correctness of one's English; marks would only be deducted for incomprehensible passages. The reports by students in selected departments with a focus on social sciences (Applied Linguistics, Sociology, and Politics) suggest a somewhat different situation: Here language seems to play a much more important role.

The lecturers mirrored students' perceptions. Lecturers from the aforementioned selected social sciences departments, while expressing very diverse views, nevertheless seemed to give language a somewhat more prominent role than lecturers in the departments with a quantitative focus. There lecturers showed a clear tendency toward arguing that language did not have an impact on assessment and grades as long as students managed to communicate their understanding and to clearly and comprehensibly make their point. Generally, lecturers in selected departments with a quantitative focus stated that it was the content and application of theory that mattered and that they would be tolerant toward linguistic issues in their marking. Marks would only be deducted if the work was not understandable, if students could not demonstrate their conceptual understanding, or if they failed to develop coherent arguments. Although English may reportedly not play a significant role in academic grading, we would nevertheless argue that expressing conceptual understanding or developing a comprehensible argument implies a certain, quite advanced level of English proficiency, below which we assume a student cannot fulfill these requirements. Turning to the more language-focused departments in the social sciences domain, lecturers here showed a huge variety of approaches to marking, ranging from being very lenient in cases where English was not the students' first language to showing a very strict attitude toward correct language use, placing importance on maintaining standards. Several lecturers in the social sciences departments mentioned that language did have an impact on marking, positive as well as negative, but that it was difficult to disentangle content and form, as they influenced each other. Quite a number of lecturers across all departments stated that language was not the decisive criterion and that the marking criterion of presentation or language carried less weight than the criteria focusing on content, analysis, and critique.

The discipline-specific perceptions of students and their academic lecturers are supported by the correlation analyses, which revealed that language test scores are not related to academic success in the selected social science departments, which could be due to the fact that students and lecturers pay more attention to it, hence students' English is more likely to improve and less likely to impact the final academic grade.

With regard to teaching and supervision, lecturers in all departments seemed to agree that the focus was on content more than on language. In line with the earlier findings, some lecturers in the selected social science departments reported that they would point out linguistic issues in one-to-one situations, so as not to embarrass students. Interestingly, most lecturers across all departments showed a high level of awareness of how to adjust their teaching and supervision to the needs of international students, mainly by adjusting speed, accent, and simplicity of English, but many reported that large

and heterogeneous student groups prevented them from adjusting their teaching style to individuals' needs. We then examined whether nonnative lecturers were more lenient, as students indicated. Indeed, within the selected social science departments, nonnative lecturers tended to make more allowances and to adjust their teaching and communication styles, but there were also native speakers who very consciously prepared handouts and accommodated for students' linguistic needs. Several native speakers commented that they would not make allowances in their teaching, supervision, and communication styles because standards needed to be kept up and students were working for a degree from a U.K. university. Within the departments with a quantitative focus, we could not find any such differences between native and nonnative speaker lecturers. Here, it seems, the focus on content overrides any language-related differences.

When we asked students whether their English proficiency had an effect on the academic grades they received, their responses varied to a certain degree, with about half the students not perceiving a large effect and the other half thinking that English did have an effect on their grades. It was difficult to compare students from the more language-oriented departments with those from the more technically focused ones because the majority of students answering this question came from the latter departments. Students generally acknowledged that the role of language for their academic success and grades ultimately depended on their lecturers' attitudes. For example, students reported that feedback they received on their linguistic performance differed widely, not only from department to department but also from lecturer to lecturer, a view supported by the lecturers' varying attitudes toward language. The majority of students perceived that they had to make more effort than a native speaker, spending more time on reading sources and on proofreading. If they spent additional time on proofreading and editing their written assignments, they felt that their language proficiency did not have a negative effect on their academic grades. Students also recognized that a higher English proficiency will most likely have a positive effect on academic performance. This perception is supported by the expectancy graphs reported previously, which revealed a trend for students with higher TOEFL iBT scores to be more likely to receive a pass or higher grade. Furthermore, the regression analyses yielded a substantial predictive power of TOEFL iBT scores on academic grades, thus also supporting students' perceptions.

In sum, students and lecturers across departments seemed to agree that good academic grades require an ability to express one's understanding and to structure a convincing argument or case, which in turn requires a certain level of English proficiency and academic writing skill. Given the findings from Strand 1, we can assume that the students coming in with TOEFL iBT scores at or above 75 overall do possess this required level of English proficiency, because the vast majority achieve a pass or an even higher academic grade. Although nonnative students may have to put more time and effort into their academic work, several participants mentioned that being a native speaker does not in itself guarantee high academic grades. We have to concede that we could not investigate the native versus nonnative speaker comparison with our data set, though it would be interesting to conduct such a study in the future.

Implications for the Field

We now address the implications for the field of HE, in particular decisions in admissions and placement in language support classes in the United Kingdom, by answering our two main aims, that is, recommendations of minimum TOEFL iBT entrance scores and recommendations for providing additional language support.

Recommendations of Minimum TOEFL iBT Entrance Scores

To address our first main aim, that is, what minimum TOEFL iBT entrance scores can be recommended for selected academic disciplines in order for students to be equipped with the necessary language skills to function in postgraduate studies, we draw on all relevant data from both strands.

Strand 1 results indicate that students, regardless of their TOEFL iBT scores, get on rather well, with the vast majority achieving their targeted degree and many of them receiving a merit or even distinction. Expectancy graphs show a tendency for students in higher TOEFL iBT subgroups to achieve higher academic grades, while students who fail or achieve lower grades are more likely to be found in lower TOEFL iBT subgroups; this trend can be found across the three largest faculties. Nevertheless, the proportion of students failing or receiving a lower degree is very small. OLR models indicate that students coming in with a hypothetical TOEFL iBT overall score of 70 have a probability of .85 to pass or do even better; to put things into perspective, the lowest score in our sample was 75. In none of our models, including those for specific departments, do we find any indication that the probability of receiving a lower degree or failing would ever get higher than the probability of achieving at least a pass. Even students attending presessional and thus coming in with the

lowest possible TOEFL iBT scores (here tested for a hypothetical low score of 65) still have a 57% probability of receiving a pass, regardless of their discipline. Given our data range, the regression results, and the requirements of U.K. Visas and Immigration for international students to come in with an English-language test equivalent to at least the level B2 of the Common European Framework of Reference (Council of Europe, 2001),[13] we have no grounds on which to recommend lowering the minimum entrance requirement.[14]

Strand 2 findings shed more light on students' and their tutors' perceptions of the appropriateness of entrance requirements. In the tutor questionnaire, we asked tutors whether they regarded the existing English-language entry requirements of their departments as too low (1), appropriate (2), or too high (3). None of the 32 tutors regarded the entry requirements as too high. The average of 1.72 (*SD* .457) indicates that tutors generally thought entry requirements were set at an appropriate, perhaps slightly too low, level.

Interview data confirm the questionnaire results: When asking tutors for the appropriateness of the required test scores, 15 mentioned that the scores were appropriate, with several tutors recommending not to lower the entry requirements, because students needed to "hit the ground running" (T25). Furthermore, concern over the university's reputation was expressed should the entry standards be lowered. Some tutors mentioned the importance of looking at the section scores, particularly writing and speaking, while others recommended making use of interviews in addition to test scores. Given our regression models, we did not find a higher predictive power for the section scores so that our data do not support these lecturers' perceptions. It is important to note that no EAP tutor or academic lecturer asked for the entrance test scores to be lowered. The vast majority of students knew about the minimum required TOEFL iBT test scores for entry to their specific courses, and no student stated that these were too low or too high.

In brief, our findings indicate that students' TOEFL iBT scores can be interpreted as language proficiency levels appropriate for the courses students are about to enter. Thus our study reconfirms the appropriateness of the currently existing entrance requirements. Our findings do not lead us to recommend lowering any of the requirements, either for the academic courses or for attending presessional.

Recommendations for Additional Language Support

Our second overarching aim was to consider what recommendations might be made with a view to placing students (with certain TOEFL iBT overall and section scores and from certain academic disciplines) into presessional and in-sessional language support programs.

Strand 1 findings show that it was students with lower TOEFL iBT scores who tended to follow presessional or in-sessional courses, though a statistically significant difference was not observed in the academic grades achieved by these students compared with those who had not obtained language support. Importantly, the regression model taking language support into account indicates that even students entering at the bottom of this score range still stand a much higher chance of passing than failing their degree or receiving a lower qualification. Overall, students entering on the basis of lower TOEFL iBT scores and taking language support did not seem disadvantaged in their final academic outcomes when compared with those entering on the basis of higher scores and not required to take language support. Taken collectively, these findings suggest no strong grounds for increasing the amount or level of language support offered to students with lower TOEFL iBT scores at this university.

Nevertheless, based on the regression results that take nationality into account, it would appear that additional support may be beneficial for students coming from linguistic and educational backgrounds that are somewhat more distant from U.K. academic culture. In this respect, drawing on insights from Strand 2, we recommend that such support should aim not so much to raise language proficiency as to facilitate learning of relevant academic skills and practices. It would appear that early academic acculturation may help offset the potential disadvantages of weaker language proficiency.

More generally, in terms of making postenrollment support more attractive to students and increasing take-up, Strand 2 findings suggest that support (in developing academic skills and literacies) may work best when embedded in the curriculum and tailored to discipline-specific discourses of academic inquiry. Clearly this would entail close collaboration between English-language tutors and academic lecturers across different departments to develop this kind of discipline-specific provision (see Murray, 2015). In this respect, a further general recommendation emerging from our Strand 2 findings is to enhance the flow of information and communication between the Centre for Applied Linguistics and academic departments to ensure follow-up where weaker presessional students are advised to attend in-sessional classes and to ensure that all students and academic staff are aware of the range of postenrollment language support provisions on offer.

Conclusions and Implications for TOEFL iBT

To conclude, we discuss implications for TOEFL iBT emerging from our study, which fills a gap in investigating the predictive validity of the TOEFL iBT test with an underresearched qualitative perspective for a context outside North America. As outlined previously, the students in our study, who entered university on the basis of a range of TOEFL iBT scores above thresholds set by individual departments, appeared to be well prepared for their studies, as indicated by their high final academic grades and by their and their tutors' perceptions of their preparedness for academic studies as far as their English-language skills were concerned. TOEFL iBT was generally regarded by our participants as predicting both language skills and academic success rather well. Equally, as discussed, we found that students coming in with lower TOEFL iBT scores and thus having to attend language support programs showed no measurable differences in their final academic grades when compared to students who did not have to attend language support classes. Thus our research findings support the following two claims:

> The test score reflects the ability of the test taker to use and understand English as it is spoken, written and heard in English-medium college and university settings. The score is useful for aiding in admissions and placement decisions, and for guiding English-language instruction. (Enright et al., 2007, p. 6)

Our research provides empirical support by showing that there are meaningful "relationships between test scores and. .. academic placements" (ETS, 2008, p. 3) as well as academic outcomes. Furthermore, within the complex situation of different departments and degree programs requiring different entrance scores, our findings support the claim that TOEFL iBT scores allow us to "discriminate between students who do or do not require additional language training" (Enright et al., 2007, p. 18). Hence our study contributes to empirically underpinning the use of TOEFL iBT scores in a U.K. setting.

To sum up, our study results support the test's predictive power with regard to academic placement and success, as demonstrated by the extremely successful cohorts of students over 3 academic years, with even students at the lower end of the TOEFL iBT score spectrum (attending presessional and/or in-sessional classes) performing very well, without measurable differences in their academic outcomes from students coming in with higher scores. In line with the quantitative findings, the interview and questionnaire results confirm the test's perception as a valid reflection of language skills and as good preparation for academic language requirements.

Acknowledgments

We would like to express our gratitude to all participating students, EAP tutors, and academic lecturers, without whom this study would not have been possible. We also would like to thank the staff in Central Registry and presessional/in-sessional administration for providing us with data. We would particularly like to thank our research assistants Samaneh Zandian, for her help during the data collection phase; Lina Shaheen, for coding the interview data; and Monae Verbeke, for her input on coding and analysis and for writing the Students' Perceptions of Their Linguistic Preparedness and Progress and the Lecturers' Perceptions of Students' Preparedness subsections. Furthermore, we would like to thank the three anonymous reviewers for their insightful comments. All remaining errors are ours.

Notes

1 Online at http://www.ets.org/research/policy_research_reports/ets
2 Secure English Language Tests as required and approved by U.K. Visas and Immigration to receive a student visa. At the time of starting the project, TOEFL iBT was approved as SELT. See https://www.gov.uk/tier-4-general-visa/knowledge-of-english
3 http://www.russellgroup.ac.uk/
4 http://www2.warwick.ac.uk/fac/soc/al/research/groups/llta/research/past_projects/strand_2_project_report_public.pdf
5 Information on TOEFL iBT entry scores set by the university is accessible online (http://www2.warwick.ac.uk/study/postgraduate/apply/english/), distinguishing requirements at four bands A – D. The band requirements for specific departments and courses can be accessed at http://www2.warwick.ac.uk/study/postgraduate/apply/english/departmentrequirements
6 In the United Kingdom, the QAA Quality Code for Higher Education oversees academic standards, assessment, and grading practices across all U.K. universities to ensure quality and standardization. All universities in the United Kingdom undergo periodic QAA institutional reviews in this regard, and the selected university's most recent QAA review was in February 2013. For more details, see http://www.qaa.ac.uk/AssuringStandardsAndQuality/quality-code/Pages/default.aspx
7 It has to be conceded that no reliable information on students' prior academic success (such as undergraduate study results) can be collected, not least due to the variety and lack of comparability of the academic systems the students attended prior to their U.K. postgraduate studies.

8 Drawing on information provided at http://www.ets.org/toefl/ibt/prepare/
9 Using information from http://www.ets.org/toefl/institutions/scores/compare and http://www.ets.org/s/toefl/newsletter/2012/19647/ukba.html; partly based on Coleman, Starfield, and Hagan (2003).
10 Informed by the staff survey at http://www2.warwick.ac.uk/fac/cross_fac/iatl/funding/fundedprojects/fellowships/grier/staff_survey
11 Student Questionnaire 1, http://www2.warwick.ac.uk/fac/soc/al/research/projects/completed/ets_project_sq1; Student Questionnaire 2, http://www2.warwick.ac.uk/fac/soc/al/research/projects/completed/ets_project_sq2; tutor questionnaire, http://www2.warwick.ac.uk/fac/soc/al/research/projects/completed/ets_project_tq1
12 A linear regression predicts the final grade as a simple linear function of the TOEFL score; final grades are coded 0, 1, 2, 3, 4, and 5. The logistic regression predicts the *probability* of the final academic grade on the basis of the TOEFL score; each grade is a distinct category, and the probabilities are derived through a nonlinear function.
13 See https://www.gov.uk/tier-4-general-visa/knowledge-of-english for English proficiency requirements.
14 See http://www2.warwick.ac.uk/study/postgraduate/apply/english/ for requirements when entering academic courses and http://www2.warwick.ac.uk/fac/soc/al/learning_english/presessional for presessional entry requirements.

References

Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London, England: Continuum.

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). Hoboken, NJ: John Wiley.

Bayliss, A., & Ingram, D. E. (2006, October). *IELTS as a predictor of academic language performance*. Paper presented at the Australian International Education Conference, Perth, Australia. Retrieved from http://aiec.idp.com/uploads/pdf/BaylissIngram%20(Paper)%20Wed%201630%20MR5.pdf

Bland, M. (2000). *An introduction to medical statistics* (3rd ed.). Oxford, England: Oxford University Press.

Bonett, D. G., & Wright, T. A. (2000). Sample size requirements for Pearson, Kendall, and Spearman correlations. *Psychometrika, 65*, 23–28.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101.

Bridgeman, B., Cho, Y., & DiPietro, S. (2015, May 13). Predicting grades from an English language assessment: The importance of peeling the onion. *Language Testing*. Retrieved from http://journals.sagepub.com/home/ltj

Bryman, A. (2008). *Social research methods* (3rd ed.). Oxford, England: Oxford University Press.

Carroll, J., & Ryan, J. (Eds.). (2005). *Teaching international students: Improving learning for all*. Abingdon, England: Routledge.

Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (Eds.). (2008). *Building a validity argument for the Test of English as a Foreign Language*. New York, NY: Routledge.

Cho, Y., & Bridgeman, B. (2012). Relationship of TOEFL iBT™ scores to academic performance: Some evidence from American universities. *Language Testing, 29*, 421–442.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th ed.). London, England: Routledge.

Coleman, D., Starfield, S., & Hagan, A. (2003). The attitudes of IELTS stakeholders: Student and staff perceptions of IELTS in Australian, UK and Chinese tertiary institutions. *IELTS Research Reports, 5*, 160–235.

Collentine, J., & Freed, B. F. (2004). Learning context and its effects on second language acquisition: Introduction. *Studies in Second Language Acquisition, 26*, 153–172.

Cotton, F., & Conrow, F. (1998). An investigation of the predictive validity of IELTS amongst a sample of international students studying at the University of Tasmania. *IELTS Research Reports, 1*, 72–115.

Council of Europe. (2001). *A common European framework of reference for language learning and teaching*. Cambridge, England: Cambridge University Press.

Creswell, J. W. (2009). *Research design: Qualitative and quantitative and mixed methods approaches*. London, England: Sage.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). London, England: Sage.

Dooey, P., & Oliver, R. (2002). An investigation into the predictive validity of the IELTS test as an indicator of future academic success. *Prospects, 17*(1), 36–54.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative and mixed methodologies*. Oxford, England: Oxford University Press.

Dörnyei, Z. (2010). *Questionnaires in second language research: Construction, administration, and processing*. London, England: Routledge.

Education Intelligence. (2014). *Integration of international students: A UK perspective*. London, England: British Council.

Educational Testing Service. (2008). *Validity evidence supporting the interpretation and use of TOEFL iBT™ scores* (TOEFL Research Insight Series Vol. 4). Retrieved from http://www.ets.org/s/toefl/pdf/toefl_ibt_insight_s1v4.pdf

Enright, M. K., Chapelle, C. A., & Jamieson, J. M. (2007, June). *From validation research to a validity argument*. Paper presented at EALTA, Sitges, Spain. Retrieved from http://www.ealta.eu.org/conference/2007/docs/pres_friday/enrightetal.pdf

Feast, V. (2002). The impact of IELTS scores on performance at university. *International Education Journal, 3*(4), 70 – 85.

Floyd, C. B. (2015). Closing the gap: International student pathways, academic performance and academic acculturation. *Journal of Academic Language & Learning, 9*(2), A1 – A18.

Fox, J. (2009). Moderating top-down policy impact and supporting EAP curricular renewal: Exploring the potential of diagnostic assessment. *Journal of English for Academic Purposes, 8*(1), 26 – 42.

French, S. (2011, July 4). Cynefin, statistics and decision analysis. *Journal of the Operational Research Society, 64*, 547 – 561. doi:https://doi.org/10.1057/jors.2012.23

Gass, S., & Mackey, A. (2007). *Data elicitation for second and foreign language research*. London, England: Routledge.

Gliner, J. A., Morgan, G. A., & Leech, N. L. (2009). *Research methods in applied settings: An integrated approach to design and analysis*. New York, NY: Routledge.

Hill, K., Storch, N., & Lynch, B. (1999). A comparison of IELTS and TOEFL as predictors of academic success. *IELTS Research Reports, 2*, 52 – 63.

Hirsh, D. (2007). English language, academic support and academic outcomes: A discussion paper. *University of Sydney Papers in TESOL, 2*(2), 193 – 211.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.

Huong, T. T. (2000). The predictive validity of the International English Language Testing System (IELTS) test. *Post-Script, 2*(1), 66 – 96.

Ingram, D., & Bayliss, A. (2007). IELTS as a predictor of academic language performance, part 1. *IELTS Research Reports, 7*, 137 – 204.

Jang, E. E., McDougall, D. E., Pollon, D., Herbert, M., & Russell, P. (2008). Integrative mixed methods data analytic strategies in research on school success in challenging circumstances. *Journal of Mixed Methods, 2*(2), 221 – 247.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17 – 64). Westport, CT: American Council on Education and Praeger.

Kerstijens, M., & Nery, C. (2000). Predictive validity in the IELTS test. *IELTS Research Reports, 3*, 85 – 108.

Klinger, C. M., & Murray, N. (2012). Tensions in higher education: Widening participation, student diversity and the challenge of academic language/literacy. *Journal of Widening Participation and Lifelong Learning, 14*(1), 27 – 44.

Kokhan, K. (2012). Investigating the possibility of using TOEFL scores for university ESL decision-making: Placement trends and effect of time lag. *Language Testing, 29*(2), 291 – 308.

Krzanowski, W. J., & Marriot, F. H. C. (1994). *Multivariate analysis: Part 1. Distributions, ordination and inference*. London, England: Arnold.

Krzanowski, W. J., & Marriot, F. H. C. (1995). *Multivariate analysis: Part 2. Classification, covariance structures and repeated measurements*. London, England: Arnold.

Kvale, S. (2007). *Doing interviews*. London, England: Sage.

Merriam, S. (2009). *Qualitative research: A guide to design and implementation*. San Francisco, CA: Jossey-Bass.

Morse, J. M. (1991). Approaches to qualitative-quantitative methodological triangulation. *Nursing Research, 40*, 120 – 123.

Murray, N. (2015). *Standards of English in higher education: Issues, challenges and strategies*. Cambridge, England: Cambridge University Press.

O'Loughlin, K. (2008). The use of IELTS for university selection in Australia: A case study. *IELTS Research Reports, 8*, 2 – 98.

Paul, A. (2007). IELTS as a predictor of academic language performances: Part 2. *IELTS Research Reports, 7*, 205 – 240.

R Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Rosenfeld, M., Leung, P., & Oltman, P. K. (2001). *The reading, writing, speaking, and listening tasks important for academic success at the undergraduate and graduate levels* (TOEFL Monograph No. 21). Princeton, NJ: Educational Testing Service.

Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology, 85*(1), 112 – 118.

Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition, 26*, 173 – 200.

Teddie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Thousand Oaks, CA: Sage.

Trahar, S. (2007). *Teaching and learning: The international higher education landscape*. Bristol, England: Escalate, the Higher Education Academy Education Subject Centre.

Tukey, J. W. (1977). *Exploratory data analysis*. Boston, MA: Addison-Wesley.

Ushioda, E., & Harsch, C. (2011). *Addressing the needs of international students with academic writing difficulties: Pilot Project 2010/11. Strand 2: Examining the predictive validity of IELTS scores* (internal report). Warwick, England: Centre for Applied Linguistics,

University of Warwick. Retrieved from http://www2.warwick.ac.uk/fac/soc/al/research/groups/llta/research/past_projects/strand_2_project_report_public.pdf

Van Nelson, C., Nelson, J. S., & Malone, B. G. (2004). Predicting success of international graduate students in an academic university. *College and University Journal, 80*(1), 19 – 27.

Wagner, C. H. (1982). Simpson's paradox in real life. *The American Statistician*, *36*(1), 46 – 48.

Wall, D., Clapham, C., & Alderson, J. C. (1994). Evaluating a placement test. *Language Testing, 11*, 321 – 344.

Wang, L., Eignor, D., & Enright, M. K. (2008). A final analysis. In C. A. Chapelle, M. K. Enright, & J. M. Jamieson (Eds.), *Building a validity argument for the Ttest of English as a Fforeign Llanguage* (pp. 259 – 318). New York, NY: Routledge.

Wiberg, M., & Sundstrom, A. (2009). A comparison of two approaches to correction of restriction of range in correlation analysis. *Practical Assessment, Research & Evaluation, 14*(5), 1 – 9.

Woodrow, L. (2006). Academic success of international postgraduate education student and the role of English proficiency. *University of Sydney Papers in TESOL, 1,* 51 – 70.

Zappa-Holman, S. (2007). Academic presentations across post-secondary contexts: The discourse socialization of non-native English speakers. *The Canadian Modern Language Review, 63*, 455 – 485.

Appendix A

Interview Guide, Students, First Interview

Overview: We will not test your English; we'll discuss the following four areas: 1. How do you get on with English? 2. Assessment procedures at your department. 3. TOEFL test reports. 4. Previous and current support with English.

1. Lead-In Question: How do you feel you are getting on, generally speaking, with the English language? How are you coping with the linguistic demands? [hand out table below]

1.1. How do you get on linguistically regarding the following aspects? Do you feel "prepared" for them?

1.2. Has the TOEFL test and/or your preparation for it prepared you for coping with them?

1.3. Do you get feedback on any of these areas (e.g., presentations, group work)?

1.4. Have you received any grades for Term 1 in any of these areas? [if applicable]

[give room to express areas where they struggle (ask for reasons and whether they get support) and areas where students **excel** (perhaps unexpectedly)]

	1.1. Getting on	1.2. Prep TOEFL	1.3. Feedback	1.4. Grades
Using English for daily life in the UK (e.g., opening bank account, finding accommodation, shopping, informal talking to classmates)				
Using English for academic life in general (e.g., talking to admissions officers, office communication)				
Using English for the following aspects of academic work:				
Listening to lectures, seminars, or tutorials				
Taking notes Reading and processing academic texts such as journals or books				
Writing assignments, reports, or essays				
Speaking in seminars or tutorials				
Preparing and giving presentations				
Group work				
Other: please specify				

2. Assessment Practices

2.1. What role does language play in your assignments/exams/drafts? Do tutors put emphasis on language?

2.2. Is there a particular assessment criterion for language? [probe for students' awareness of what they are assessed against]

3. TOEFL Test Reports [hand out TOEFL report booklet for them to keep]

3.1. Have you seen these test reports? Did you use them? Why/what for/why not?

3.2. Were you aware of the university's entry requirements with regard to the TOEFL test scores?

3.3. [explain CEFR/relation to TOEFL alignment; hand out CEFR self-assessment grid A3; give time to fill in] How useful do you find the self-assessment grid? Would you use it to plan your language learning?

3.4. What do you think a "good" test report should look like? [probe for what descriptions students would find most helpful, along with the test scores]

4. Current Support With English Language Learning

4.1. Did your test results influence your decision to seek language support?

4.2. [follow-up presessional comments if applicable; probe whether presessional is perceived as helpful preparation and why] How have these classes helped you with your English (e.g., listening and taking notes in lectures, reading, writing your assignments)?

4.3. Ask the following questions again [from Student Questionnaire 1, Q14–16; be flexible]

[14.] Are you currently attending in-sessional classes? If yes, which class(es) and how often do you go (hours/week)? How are these classes helping you with your English (e.g., listening and taking notes in lectures, reading, writing your assignments)? Was attending in-sessional classes recommended to you? If yes, by whom?

[15.] Are you currently attending any other language support classes (either at your department, at the university, or via private tuition)? If yes, which ones, and for how many hours/week? How are these classes helping you with your English (e.g., listening and taking notes in lectures, reading, writing your assignments)?

[16.] Are you actively improving your English skills at present? If so, how, and how many hours/week? How are these activities helping you with your English (e.g., listening and taking notes in lectures, reading, writing your assignments)?

5. Further Comments [probe for any additional comments or questions]

6. Are you happy for us to get in touch with your EAP tutors/academic lecturers? We simply would like to ask their perception of your linguistic preparedness and how you cope linguistically. We guarantee that we will **only** discuss linguistic aspects and nothing else. You are very welcome to be present during that interview if you wish.**7.** [**thank student** and invite to participate in **questionnaire** if student has not done so yet]

Appendix B

Interview Guide, Students, Second Interview

Overview: We will not test your English; we'll discuss the following areas: 1. How you are coping with the linguistic demands of your academic studies? 2. Link between your TOEFL score and academic assignment grades. 3. Link between language support program and academic assignment grades. 4. Current support with English language.

1. Lead-In Question: How are you coping with the linguistic demands of academic studies since our last interview?

1.1. [hand out table below] How do you get on linguistically regarding the following aspects since our last interview? Do you feel more "prepared" for them?

1.2. Have you got feedback on any of these areas in Term 2 (e.g., presentations, group work)? If yes, is the feedback different from what you received in the beginning of the year? If so, in what terms is it different?

1.3. Have you received any grades for Term 2 in any of these areas? Would you like to share some of them with us?

	1.1. Getting on	1.2. Feedback	1.3. Grades
Using English for daily life in the UK (e.g., socializing, obtaining goods and services, traveling in the UK)			
Using English for academic life in general (e.g., talking to admissions officers, office communication)			
Using English for the following aspects of academic work:			
Listening to lectures, seminars, or tutorials			
Taking notes			
Reading and processing academic texts			
Writing assignments, reports, or essays			
Speaking in seminars or tutorials			
Preparing and giving presentations			
Group work			
Other: please specify			

2. Link Between Your TOEFL Score and Academic Assignment Grades

2.1. Do you think your English proficiency has influenced your academic assignment grades?

2.2. [show student his or her TOEFL score, in additional document] Do you think your academic assignment grades reflect your TOEFL score profile (in relation to listening, speaking, reading, writing)? Are these academic grades more or less what you expected, based on your TOEFL score?

3. Link Between Language Support Program and Academic Assignment Grades [only if attended presessional]

3.1. How do you think presessional programs helped you with your English (e.g., listening and taking notes in lectures, reading, writing your assignments)?

3.2. Do you think that attending the presessional program has had an influence on your academic grades?

4. Current Support With English Language [ask the following questions again (Quaire1 Q14–16); be flexible]

[14] Did you attend in-sessional classes in Term 2? If yes, which classes and how often did you go? How did these classes help you with your English? Was attending in-sessional classes last term recommended to you? If yes, by whom? Did your TOEFL test results or your academic grades influence your decision to seek language support?

[15] Are you currently attending any other language support classes (either at your department, at the university, or via private tuition)? If yes, which ones, and for how many hours/week? How are these classes helping you with your English (e.g., listening and taking notes in lectures, reading, writing your assignments)?

[16] Are you actively improving your English skills at present? If so, how, and how many hours/week? How are these activities helping you with your English (e.g., listening and taking notes in lectures, reading, writing)?

5. Further Comments [invite students to comment on any other relevant aspects]

6. Contact Tutors Who Teach You, Marked Your Assignments. Are you happy for us to get in touch with your tutors/supervisors (EAP tutors and/or academic lecturers)? We simply would like to ask their perception of your linguistic preparedness and how you cope linguistically. We guarantee that we will *only* discuss linguistic aspects and nothing else. You are very welcome to be present during that interview if you wish. Would you please give us the names of the tutors/supervisors you are happy for us to contact?

7. Student ID. We will keep all data anonymous but would like to link student interview data to the big anonymous data set we received from registry; for this purpose, we would need your student ID.

Thank you for your participation!

Appendix C

Interview Guide, Students, Third Interview

Looking back over the past year, we'd like to reflect with you on your linguistic preparedness for your studies. Let's start with having a look at how well you think the TOEFL test prepared you.

1. TOEFL as Preparation for Academic Studies [hand out table and probe for each skill]

1.1. How well do you think the TOEFL test prepared you for your academic studies?

1.2. How well do you think the TOEFL test prepared you for life in the UK?

1.3. How well do you think your TOEFL score profile is reflected in your academic performance in your studies?

Listening	☐Very well	☐Quite OK	☐A little bit	☐Not at all well
Speaking	☐Very well	☐Quite OK	☐A little bit	☐Not at all well
Reading	☐Very well	☐Quite OK	☐A little bit	☐Not at all well
Writing	☐Very well	☐Quite OK	☐A little bit	☐Not at all well

1.4. Are there any aspects missing in the TOEFL test which you needed here in the UK for your studies?

2. Your English-Language Skills

2.1. Do you think your English was good enough to cope with the academic demands in the UK?

2.2. Do you think your English was good enough to cope with everyday life in the UK?

2.3. Do you think you improved your English during the last academic year?

2.4. How well did you get on during your studies in the UK with regard to your language skills? [hand out table] Please rate the following aspects on a scale from 1 (*not well at all*) to 5 (*very well*).

Please mark only one box per row	Not well at all				Very well
Using English for daily life in the UK (e.g., opening bank account, finding accommodation, informal talking to classmates)	①	②	③	④	⑤
Using English for academic life in general (e.g., talking to admissions officers, office communication)	①	②	③	④	⑤
Using English for the following aspects of academic work:					
Listening to lectures, seminars, or tutorials	①	②	③	④	⑤
Taking notes	①	②	③	④	⑤
Reading and processing academic texts such as journals or books	①	②	③	④	⑤
Writing assignments, reports, or essays	①	②	③	④	⑤
Speaking in seminars or tutorials	①	②	③	④	⑤
Preparing and giving presentations	①	②	③	④	⑤
Group work	①	②	③	④	⑤
Other: please specify	①	②	③	④	⑤

2.5. Did your language proficiency affect your academic progress? If so, in which ways?

2.6. Self-assessment of your English-language skills [hand out CEFR self-assessment grid to students] The following statements are taken from a widely recognized proficiency framework. Please tick for each of the four skills the statement which best describes your language skills.

2.7. Looking back, would you prepare differently for your studies in the UK (a) with regard to the TOEFL test? If yes, in which ways? (b) for your academic studies? If yes, in which ways? (c) for social life? If yes, in which ways?

3. Exploitation of Language Support

3.1. Were you required to attend presessional classes? If so, how have these classes helped you with your English (e.g., listening and taking notes in lectures, reading, writing your assignments)?

3.2. Did you attend in-sessional classes during the last academic year? If so, please list the class(es) and how often you went (hours/week). How did these classes help you with your English (e.g., listening and taking notes in lectures, reading, writing your assignments)?

3.3. Was attending in-sessional classes recommended to you? If yes, by whom?

3.4. Did you attend any other language support classes (either at your department, at the university, or via private tuition)? If so, which ones, and for how many hours/week? How did these classes help you with your English (e.g., listening and taking notes in lectures, reading, writing your assignments)?

3.5. Did you actively improve your English skills during the last academic year? If so, please state how, and how many hours/week. How did these activities help you with your English (e.g., listening and taking notes in lectures, reading, writing your assignments)?

4. Further Comments. We very much appreciate your comments and feedback on our research.

5. Further Contact. If you would like to receive a research report in due course, please provide an e-mail address which will be valid in the future. We will only use the address to contact you for the report.

Thank you very much for your participation, and all the best for your future!

Appendix D

Interview Guide, EAP Tutors

Overview: We will discuss the following four areas: 1. How do you think your student was getting on with the English language? 2. Assessment procedures in pre−/in-sessional course. 3. TOEFL test reports. 4. Further language support.

1. How Did the Student Get On With His or Her English Language?

1.1. How do you feel the student was getting on during pre−/in-sessional, generally speaking, with the English language? How was he or she coping with the linguistic demands?

1.2. We are interested in your perception of difficulties encountered by the student. Are the following aspects covered in pre−/in-sessional, and if so, did the student have any difficulties there? [hand this table out]

Using English for daily life in the UK (e.g., opening bank account, shopping, informal talking to classmates)
Using English for academic life in general (e.g., talking to admissions officers, office communication)
Using English for the following aspects of academic work:
  Listening to lectures, seminars, or tutorials
  Taking notes
  Reading and processing academic texts such as journals or books
  Writing assignments, reports, or essays
  Speaking in seminars or tutorials
  Preparing and giving presentations
  Group work
Other: please specify

1.3. We are interested in your perception of difficulties the student may encounter during his or her academic studies. Could you indicate/speculate on how you think the student might now be getting on with his or her academic studies and what difficulties you anticipate he or she might be experiencing?

2. Assessment Procedures During the Presessional/In-sessional Courses

2.1. Did you give **feedback** on any of the areas outlined in the table above (e.g., presentations, group work)?

2.2. Did you **formally assess** any of these areas? Were there any concerns emerging for the student?

2.3. Generally speaking, what assessment approaches do you use? What aspects do you assess (also with regard to the above table)? Do you give formative feedback and/or summative tests?

3. TOEFL Test Reports

3.1. Do you usually get information about students' language scores, for example, TOEFL test reports, presessional or in-sessional reports?

3.2. Based on your experience, how useful do you find the TOEFL score reports and the TOEFL-specific descriptions? What use do you make of them? [hand out TOEFL booklet for them to keep]

3.3. Would you welcome feedback on TOEFL test results that uses descriptions from the Common European Framework of Reference? Please give your thoughts on why/why not. [hand out CEFR self-assessment grid A3]

3.4. Would you be able to place your student at one of the CEFR levels in the grid?

3.5. What do you think that a "good" TOEFL test report should look like? [probe for what descriptions tutors would find most helpful, along with the test scores]

3.6. How useful do you find the presessional reports? Can you think of ways in which these reports might be changed or improved? For in-sessional, would you consider reports useful? What should they ideally look like?

4. Departmental Language Support for International Students

4.1. Do you think the student would benefit from attending any language support classes? Have you recommended any support to the student?

4.2. Are you aware of any of these support systems at Warwick? [hand out the following list]

a. Induction on academic writing, referencing, and plagiarism
b. Induction on presentations
c. Induction on group work
d. Trial assignments
e. Trial presentations
f. Mock exams
g. Explicit tutor feedback on linguistic issues for written assignments
h. Explicit tutor feedback on linguistic issues for presentations
i. Online or self-study materials
j. Study groups
k. Linguistic support built in academic modules
l. Support provided by another department or section of the university, or by outside experts, tailored for your department
m.Other: please specify

4.3. Do you know whether the student exploits any of these support systems?

4.4. Please give your thoughts why these are often/less often exploited.

5. Do you have any further comments or suggestions for us?

6. Thank the tutor and invite to our questionnaire: We will e-mail the link to you.

7. Ask for consent to share main comments with the student: Are you happy for us to share some of your main comments with the student when we next interview him or her?

Appendix E

Interview Guide, Academic Lecturers, First Interview

Overview: We will discuss the following four areas: 1. How do you think your student is getting on with language requirements? 2. Assessment procedures at your department. 3. TOEFL test reports. 4. Department language support.

1. Coping With the English Language

1.1. How do you feel your student is getting on, generally speaking, with the English language? How is he or she coping with the linguistic demands?

1.2. We are interested in your perception of difficulties encountered by the student. Could you indicate any areas where the student is not well prepared for the linguistic demands? [hand out following table]

	1.1. Getting on	1.2. Feedback	1.3. Grades
Using English for daily life in the UK (e.g., opening bank account, shopping, informal talking to classmates)			
Using English for academic life in general (e.g., talking to admissions officers, office communication)			
Using English for the following aspects of academic work:			
Listening to lectures, seminars, or tutorials			
Taking notes			
Reading and processing academic texts such as journals or books			
Writing assignments, reports, or essays			
Speaking in seminars or tutorials			
Preparing and giving presentations			
Group work			
Other: please specify			

2. Assessment Procedures at Your Department [going through the table above: feedback and grades]

2.1. Do you give **feedback** on any of the areas outlined in the table above (e.g., presentations, group work)?

2.2. Does your student's English proficiency influence your teaching and assessment? If so, in what ways? What is your attitude to the student's English proficiency? What possible impact has this on your assessment and teaching? [probe for which aspects in the table above the tutor regards English as important]

2.3. Is English proficiency reflected in your department's **assessment criteria**? If so, how and what weighting does it have? [probe which areas in the above table are formally assessed and graded]

3. TOEFL test reports

3.1. Do you usually get information about students' language scores, for example, TOEFL test reports, presessional or in-sessional reports?

3.2. Based on your experience, how useful do you find the TOEFL score reports and the TOEFL-specific descriptions? What use do you make of them? [hand out TOEFL booklet for them to keep]

3.3. Would you welcome feedback on TOEFL test results which uses descriptions from the Common European Framework of Reference? Please give your thoughts on why/why not. [hand out CEFR self-assessment grid A3]

3.4. Would you be able to place your student at one of the CEFR levels in the grid?

3.5. What do you think that a "good" TOEFL test report should look like? [probe for what descriptions tutors would find most helpful, along with the test scores]

3.6. How useful do you find the presessional/in-sessional reports? What use do you make of them?

4. Departmental Language Support for International Students

4.1. What language support is in place for (international) postgraduate/PhD students at your department? [hand out the following list]

a. Induction on academic writing, referencing, and plagiarism
b. Induction on presentations
c. Induction on group work
e. Trial assignments
f. Trial presentations
g. Mock exams
h. Explicit tutor feedback on linguistic issues for written assignments
i. Explicit tutor feedback on linguistic issues for presentations
j. Online or self-study materials
k. Study groups
l. Linguistic support built in academic modules
m. Support provided by another department or section of the university, or by outside experts, tailored for your department
n. Other: please specify

4.2. Do you think the student would benefit from attending any language support classes? Have you recommended any support?

4.3. Do you know whether the student exploits any of these support systems? Please give your thoughts why these are often/less often exploited.

5. Do you have any further comments or suggestions for us?

6. Thank the lecturer and invite to our questionnaire: We will e-mail the link to you.

7. Ask for consent to share main comments with the student: Are you happy for us to share some of your main comments with the student when we next interview him or her?

Appendix F

Interview Guide, Academic Lecturers, Second Interview

1. How was your student getting on with the linguistic requirements of the academic course/modules during Terms 2 and/or 3?

2. Do you think your student was well prepared for the following aspects, or were there any areas where he or she struggled/needed support? Please tick/comment as you see fit:

How did your student get on linguistically regarding the following aspects?	Gets on, well prepared	Has difficulties	Has improved over the year
Using English for daily life in the UK (e.g., socializing, obtaining goods and services, traveling in the UK)			
Using English for academic life in general (e.g., talking to admissions officers, office communication)			
Using English for the following aspects of academic work:			
Listening to lectures, seminars, or tutorials			
Taking notes			
Reading and processing academic texts such as journals or books			
Writing assignments, reports, or essays			
Speaking in seminars or tutorials			
Preparing and giving presentations			
Group work			
Other: please specify			

3. Would you like to share any further comments on your student's linguistic coping and preparedness with us?

Thank you for your time and participation.

Appendix G

Interview Guide, Dissertation Supervisors, Third Interview

Overview: We will discuss the following four areas: 1. How well is your student prepared for/does he or she cope with the academic requirements with regard to the English language? 2. Assessment procedures for dissertations at your department. 3. TOEFL test reports and relation to student's academic achievements. 4. Language support.

1. Coping With English Language

1.1. How do you feel the student was getting on, generally speaking, with the English language? How was he or she coping with the linguistic demands of writing a dissertation?

1.2. We are interested in your perception of difficulties encountered by the student. Could you indicate any areas where the student was not well prepared for the linguistic demands? [hand out following table]

	Difficulties	Some issues, but could overcome	No problem, well prepared
Using English for the following aspects of the dissertation research and writing process:			
Listening in supervision tutorials, group seminars			
Taking notes during supervision tutorials or seminars			
Reading and processing academic texts such as journals or books			
Undertaking a literature review, that is, processing reading input and presenting it in written form			
Writing a proposal			
Writing an abstract			
Writing drafts, e.g., producing a coherent macro- and microstructure, expressing thoughts concisely			
Revising drafts, e.g., taking feedback on board, self-evaluating			
Speaking and interacting in supervision seminars or tutorials			
Preparing and giving presentations, e.g., for supervision seminars, tutorials, or vivas			
Group work related to dissertation writing			
Using English for daily life in the UK (e.g., shopping, socializing, informal talking to classmates)			
Using English for academic life in general (e.g., talking to admissions officers, office communication)			
Other: please specify			

1.3. We are interested in your perception of your student's language proficiency, based on your experience in supervising the student. We'd like to use the assessment grid from the Common European Framework of Reference. [hand out CEFR self-assessment grid A3 for them to keep] Would you be able to place your student at one of the CEFR levels in the grid?

2. Assessment Procedures for Dissertations at Your Department [going through the table above]

2.1. Do you give **feedback** on any of the areas outlined in the table above during the preparation of the dissertation (e.g., drafts, presentations, group seminars)?

2.2. Does your student's English proficiency influence your dissertation supervision? If so, in what ways? What is your attitude to the student's English proficiency? What possible impact has this on your supervision and dissertation assessment? [probe for which aspects in the table above the tutor regards English as important]

2.3. Is English proficiency reflected in your department's dissertation assessment criteria? If so, how and what weighting does it have? [probe which areas in the above table are formally assessed and graded]

3. TOEFL Test

3.1. Do you usually get information about students' language scores, for example, TOEFL test reports, presessional or in-sessional reports? We may have asked this before if we interviewed you before—just a brief reminder here.

3.2. Do you make use of these reports when you supervise students? What use do you make?

3.3. How useful do you find the TOEFL/pre−/in-sessional reports with regard to dissertation supervision?

3.4. TOEFL tests four skills, and you can get a maximum of 30 points for each skill. [show the student's TOEFL score profile] These are your student's TOEFL scores used for university entry. How well do you think these scores "predict" and reflect your student's academic progress and achievement?

4. Departmental Language Support for International Students

4.1. What language support is in place for the dissertation writing stage at your department? [hand out list]

a. Input on academic writing, referencing and plagiarism
b. Input on writing proposals
c. Input on planning a research project
d. Input on article critiquing
f. Input on writing a literature review
e. Input on research methods
g. Explicit tutor feedback on linguistic issues for written drafts
h. Explicit tutor feedback on linguistic issues for presentations
i. Online or self-study materials
j. Study groups
k. Input on preparing for a viva
l. Support provided by another department or section of the university, or by outside experts, tailored for your department
m. Other: please specify

4.2. Do you think your student would have benefited from attending any language support classes, either during the last year or during the dissertation writing period? Have you recommended any support during your supervision?

4.3. Do you know whether the student has exploited any support systems? Please give your thoughts whether you think the student benefited.

5. Do you have any further comments or suggestions for us?

6. [invite supervisor to our questionnaire]: We will e-mail the link to you.

Thank you for your time and participation.

Appendix H
Overview of Research Questions and Corresponding Variables From Instruments for Students

RQ theme	Categories, indicators for themes	Quantitative variables						Qualitative variables				
		Uni data	SQ1	SQ2	SI1	SI2	SI3	SQ1	SQ2	SI1	SI2	SI3
RQ3. Preparedness, perception of role TOEFL plays in preparation, predictive validity	Linguistic preparedness; getting on with English		SQ1_7 SQ1_8 SQ1_9 SQ1_10	SQ2_6 SQ2_7 SQ2_8 SQ2_9					SQ2_9.10	SI1_1.1	SI2_1.1	SI3_5 SI3_6 SI3_7 SI3_8
	How well TOEFL prepared linguistically for studies		SQ1_10	SQ2_3.3				SQ1_10.10 SQ1_10.11 SQ1_10.12		SI1_1.2		SI3_1
	How well TOEFL prepared for social life			SQ2_3.4						SI1_1.2		SI3_2
	Missing aspects in TOEFL								SQ2_5			SI3_4
	TOEFL as indicator of language skills		SQ1_11									
	TOEFL as indicator of academic grades/ success			SQ2_4							SI2_2.2	SI3_3
	TOEFL scores	SITS										
	TOEFL date	SITS										
	Preparation for TOEFL		SQ1_4					SQ1_4.4 SQ1_4.5	SQ2_12.1			SI3_11.1
	Preparation for academic studies		SQ1_5					SQ1_5open	SQ2_12.2			SI3_11.2
	Preparation for social life		SQ1_6					SQ1_6open	SQ2_12.3			SI3_11.3
	English proficiency effect on academic progress			SQ2_10								SI3_9
	Academic grades				SI1_1.4	SI2_1.3						
	Self-assessment		SQ1_12	SQ2_11			SI3_10					
	Appropriateness SELT											
RQ4. Exploitation of language support; indicators of attitudes toward improving English	Presessional	attended	SQ1_13	SQ2_13				SQ1_13.3	SQ2_13.3	SI1_4.2	SI2_3	SI3_12
	In-sessional	attended	SQ1_14	SQ2_14				SQ1_14.1	SQ2_14.1	SI1_4.3	SI2_4.1	SI3_13
	English-language support		SQ1_15	SQ2_15				SQ1_15.1 SQ1_15.2	SQ2_15.1 SQ1_15.2	SI1_4.4	SI2_4.2	SI3_14
	Active learning/ improvement		SQ1_16	SQ2_16				SQ1_16.1 SQ1_16.2	SQ2_16.1 SQ1_16.2	SI1_4.5	SI2_4.3	SI3_15
	Receiving feedback									SI1_1.3	SI2_1.2	
	Usefulness of self-assessment									SI1_3.3		

Appendix H
Continued

RQ theme	Categories, indicators for themes	Quantitative variables						Qualitative variables					
		Uni data	SQ1	SQ2	SI1	SI2	SI3	SQ1	SQ2	SI1	SI2	SI3
Test results and seeking support	Test reports: how they are used									SI1_3.1		
	Test results as impulse to seek support									SI1_4.1		
University support	Offers											
	Exploitation											
	Effectiveness											
RQ5. What role does language play in academic progress?	Role of language (drafts assignments/exams)									SI1_2.1	SI2_2.1	
	Tutors' attitude to language									SI1_2.1		
	Assessment criterion for language									SI1_2.2		

Note. Uni data refers to the data provided by the university. SITS = data from Central Registry. SQ1 = student questionnaire 1. SI1 = Student Interview 1, etc.

Appendix I
Overview of Research Questions and Corresponding Variables From Instruments for Tutors

		Qualitative variables				Quantitative variables			
RQ theme	Categories, indicators for themes	TI1	TI2	TI3	TQ	TI1	TI2	TI3	TQ
RQ3. Preparedness, perception of role TOEFL plays in preparation, predictive validity	Linguistic preparedness; getting on with English	TI1_1.1 TI1_1.2 TI1_4.2	TI2_1 TI2_2 TI2_3	TI3_1.1 TI3_1.2 TI1_4.2	TQ_10.8				TQ_10
	How well TOEFL prepared linguistically for studies				TQ_13.10				TQ13
	How well TOEFL prepared for social life				TQ_13.11				TQ13
	Missing aspects in TOEFL				TQ_13.12				
	TOEFL as indicator of language skills								TQ_14
	TOEFL as indicator of academic grades/ success			TI3_3.4	TQ_15.2				TQ_15.1
	TOEFL scores								
	TOEFL date								
	Preparation for TOEFL								
	Preparation for academic studies								
	Preparation for social life								
	English proficiency effect on academic progress								
	Academic grades								
	Self-assessment					TI1_3.4		TI3_1.3	
	Appropriateness of SELT								TQ_17
RQ4. Exploitation of language support; indicators of attitudes toward improving English	Presessional								
	In-sessional								
	English-language support	TI1_4.3		TI3_4.3					
	Active learning / improvement								
	Receiving feedback	TI1_2.1		TI3_2.1					
	Usefulness of self-assessment								
Test results and seeking support	Test reports: how are they used?	TI1_3.1 – 3.3; 3.5; 3.6		TI3_3.1 – 3.3	TQ_16.3 – 5				TQ_16.1 TQ_16.2
	Test results as impulse to seek support								
University support	Offers	TI1_4.1		TI3_4.1	TQ_6.1 TQ_6.2				
	Exploitation	TI1_4.3		TI3_4.3	TQ_7				
	Effectiveness				TQ_8 TQ_9				
RQ5. What role does language play in academic progress?	Role of language (drafts assignments/exams)	TI1_2.3		TI3_2.3	TQ_11				
	Tutors' attitude to language	TI1_2.2		TI3_2.2	TQ_11				
	Assessment criterion for language	TI1_2.3		TI3_2.3	TQ_12				

Note. Uni data refers to the data provided by the university. SITS = data from Central Registry. TI1 = Tutor Interview 1, etc. TQ = tutor questionnaire.

Appendix J
Final Coding Scheme

RQ theme	NVivo parent nodes: Category	NVivo child nodes: Code	Definition
Theme 1 RQ3: Preparedness and perception of TOEFL	1. Linguistic preparedness; getting on with English	1. Well prepared	instances where students/their tutors feel that students are prepared, do not have any struggle; indicators for why they get on well; reasons for why they think they get on well
		2. Struggling, challenges	instances where students/tutors feel that students struggle, have difficulties, find English challenging; indicators for why they think they're struggling; reasons for why they think they have difficulties or find language challenging, dissatisfaction with English skills or lack of improvement
		3. Adjustment in the beginning	initial struggles which could be overcome rather quickly (i.e., within the first 1–3 months), reasons
		4. Improvement over time	areas where students improved over time (i.e., over half a year or longer), reasons why
		5. Little/no improvement over time	areas where students/tutors perceive little or no improvement (since Interview 1), reasons why
		6. Changing skill needs/ demands	dealing with changes in language skill needs/demands as year progresses (e.g., more emphasis on writing, critical reading, professional communication)
		7. Feeling prepared for dissertation stage	student's/tutors' perceptions how well they feel prepared for writing the dissertation
	2. How well TOEFL prepared linguistically for studies	1. Test prepared well	academic areas for which test prepared well; reasons why student thinks test prepared well
		2. Test did not prepare well	academic areas for which test did not prepare well; reasons why student thinks test prepared not well (all aspects related to content and construct validity, e.g., test essays do not reflect academic reality)
		3. Limited preparation	mixed views, limited value of preparing for studies
	3. How well TOEFL prepared linguistically for social life	1. Test prepared well	areas of social life for which test prepared well; reasons why student thinks test prepared well
		2. Test did not prepare well	areas of social life for which test did not prepare well; reasons why student thinks test prepared not well
		3. Limited preparation	mixed views, limited value of preparing for studies
	4. General evaluation of TOEFL	1. Missing aspects, problematic issues, suggestions to improve the test	general reliability and administration issues (e.g., time pressure; coverage; computer administration); suggestions about what could be improved
		2. Not an issue	instances where students/tutors explicitly say what is not an issue
		3. Positive aspects	code what students/tutors find good and helpful about TOEFL
	5. TOEFL as indicator of language skills	1. Effective indicator	areas where TOEFL is an effective indicator, i.e., reflects students' actual language skills, and reasons
		2. Not an effective indicator	areas where TOEFL is not an effective indicator, and reasons

Appendix J
Continued

RQ theme	NVivo parent nodes: Category	NVivo child nodes: Code	Definition
	6. TOEFL as indicator of academic grades/success/ performance	1. Effective indicator	reasons why TOEFL is an effective indicator for academic grades/success, that is, reflects their grades/success/performance
		2. Not an effective indicator	reasons why TOEFL is not an effective indicator of grades
		3. Expectations	students'/tutors' expectations/perceptions about the relationship (if any) between TOEFL scores and academic performance through the year
	7. Preparation for TOEFL	1. Preparation for TOEFL	code what students did; code here also whether they found it useful or not useful
		2. Did not prepare	code instances where students report they did not prepare
		3. Prep differently for test	code answers to "would you prepare differently for the test," regardless of whether student would do the same or something different
	8. Preparation for academic studies	1. Actual prep for academic studies	code all instances of preparation for their academic subject students did before they came to university (both useful and not so useful ones)
		2. Prep differently for academic studies	code answers to "would you prepare differently for your academic studies," regardless of whether student would do the same or something different
	9. Preparation for social life	1. Actual prep for social life	code all instances of preparation students did before they came to university (both useful and not so useful ones)
		2. Prep differently for social life	code answers to "would you prepare differently for social life," regardless of whether student would do the same or something different
	10. SELT entry requirements		code whether students/tutors are aware of SELT requirements, and how they perceive the appropriateness of the SELT entrance scores required by the university/department
	11. Tutor assessment of individual student	1. Tutor CEFR assessment	code what tutors say when they are assessing their student with the CEFR grid
		2. Tutor recommendations of language support	code whether tutor recommended language support for their student
Theme 2, RQ4: Exploitation of language support; indicators of attitudes toward improving English	1. Presessional	1. Effectiveness	what aspects did students perceive as effective and helpful, and why
		2. Ineffective aspects	what aspects did students perceive as not very effective and helpful, and why what could be improved
		3. Reasons for (non) attendance	what reasons did student state for attending or not attending presessional
		4. Tutors' perceptions	all instances where tutor talks about presessional
	2. In-sessional	1. Effectiveness	what aspects did students perceive as effective and helpful, and why
		2. Ineffective aspects	what aspects did students perceive as not very effective and helpful, and why what could be improved
		3. Reasons for (non) attendance	what reasons did student state for attending (or not attending) in-sessional, e.g., was it recommended, or did they not know about it
		4. Tutors' perceptions	all instances where tutor talks about in-sessional

Appendix J
continued

RQ theme	NVivo parent nodes: Category	NVivo child nodes: Code	Definition
	3. University language support	1. Offers	code what university support students/tutors mention and are aware of
		2. Exploitation	code what university support students actually make use of and why
		3. Effectiveness	perceived usefulness and reasons; also code if students/tutors do not find the offers useful, and reasons
		4. Not using it	code instances and reasons why students are not using it
		5. Expectations	what students/tutors (would) expect from a language support class
	4. Active learning/ improvement (by student)	1. Yes	instances where student is actively learning English or actively seeking help (e.g., actively seeking feedback from flat mates), and reasons why
		2. Awareness of need to improve	Student's awareness of what to improve, on which areas to work on
		3. No active learning	instances where student does not actively learn English, and reasons why not
	5. Facilitators and constraints		code all instances where student/tutor mentions facilitative factors (e.g., living with native speaker flat mates); code all constraining factors such as no time, or no contact to native speakers
	6. Making use of tutor feedback		code where student uses (or does not use) feedback from tutors on their drafts, assignments, and reasons why/why not; code here if student actively seeks feedback from tutors; code tutors' perceptions of how students use their feedback
	7. English-language support outside university (attending courses)	1. Yes	instances where student got additional support (courses, classes outside the university), and reasons why
		2. No	instances where student did not get further support, and reasons why not
	8. Usefulness of self-assessment	1. Useful	code for what purposes/areas and why students/tutors find self-assessment useful
		2. Not useful	code for what purposes/areas and why students/tutors find it less helpful
		3. Mixed views	code instances where students/tutors have mixed or no views on helpfulness of self-assessment
	9. TOEFL test results and seeking support	1. Use of test reports	code what students and tutors say that they actually do (or do not do) with the reports; do they know the booklet
		2. Test results as impulse to seek support	code instances where students and tutors seek (recommend) support based on actual test results (also linked to use of reports)
		3. Ideal report	code what students and tutors ideally want from a test report and what they would do with it; also code constraints the participants are aware of
Theme 3, RQ5: What role does language play in academic progress?	1. Students' perception of role of language in drafts, assignments, exams		all instances where *students* talk generally about the role English plays in assignments, exams, drafts, etc.; instances which clarify the focus (e.g., focus on language vs. focus on content; focus on clarity of ideas vs. correct English); instances which indicate a threshold of minimally acceptable English as students perceive it

Appendix J
continued

RQ theme	NVivo parent nodes: Category	NVivo child nodes: Code	Definition
	2. Tutors' attitude to and emphasis of language (teaching, supervision, etc.)	1. Student's perception of tutors' attitude to language	students' perception of their tutors' attitude toward English and the importance tutors assign to English; students' perception of tutors' leniency/strictness
		2. Tutor's attitude toward/ perception of language	tutor's attitudes toward language, emphasis and importance of correct English, effects on teaching, tutoring, supervision; tutor's perception of student's language level, perceived issues with student's language, indications of minimum threshold of language as tutor sees it.
	3. Feedback from tutors on assignments, drafts (table)	1. Focus	focus on content or on language (and on which areas of language)
		2. Quality, usefulness	students' and tutors' perceptions of quality, amount, usefulness of feedback indicators of students' level of satisfaction with feedback
		3. Issues with feedback	what is lacking, what could be improved; indicators for students' dissatisfaction with feedback
		4. No feedback	code all instances where students have not received any feedback
	4. Assessment criterion for language	1. Awareness, attitudes	aware that such criteria exist; attitudes expressed toward it; views on weighting of language in relation to other criteria
		2. Not aware of it	not aware that such criteria exist
	5. Assessment practice		code all information about assessment points, means, practices (e.g., viva or PMA)
	6. Student's academic progress	1. Effect of English proficiency on progress	code students' and tutors perception of whether the student's E proficiency hindered or enhanced their academic progress[a]
		2. Academic grades	code instances where students/tutors talk about academic grades

[a] All instances of influence of discipline or lecturers' attitudes are coded in 3.1 or 3.2.

Appendix K
Overview of Interview Participants (25 Student Participants and Their Tutors)

Project ID	Questionnaire 1	Questionnaire 2	Interview 1	Interview 2	Interview 3	Presess. (hours)	In-sess. (hours)	Tutor I1	Tutor I2	Supervisor I3	Department	Nationality	TOEFL overall	TOEFL Listening	TOEFL Speaking	TOEFL Reading	TOEFL Writing	TOEFL date taken	Final academic grade	Indicator for preparedness
S002	y	y	y	n	y	n	n	T19			Statistics	Italy	104	28	23	28	25	Jul 13	With distinction	4.22
S004	y	n	y	y	y	n	n	T21; T60		T61	Manufacturing	France	109	26	26	30	27	Dec 12	Pass	4.67
S005	y	y	y	y	y	n	n	T24; T25; T26	T63	T62	Business School	Italy	114	30	27	30	27	Feb 13	With merit	4.67
S006	y	n	y	y	y	n	n				Business School	Thailand	112	30	26	28	28	Sep 12	Pass	4.67
S008	y	y	y	y	y	n	9	T28			Sociology	Mongolia	99	25	24	28	22	Nov 12	Pass	3.78
S009	y	y	y	y	y	n	n			T27	Business School	China	113	27	27	29	30	Feb 12	With distinction	5.00
S011	y	n	y	y	y	n	n		T30		Engineering	Indonesia	96	27	24	24	21	May 13	With distinction	4.00
S012	y	y	y	y	y	n	n	T31; T32		T65	PAIS	Italy	101	28	26	25	22	Nov 12	With merit	4.33
S013	y	n	y	y	y	n	n			T66	Mathematics	Mexico	100	24	24	27	25	Sep 12	With merit	3.67
S014	y	n	y	n	y	n	7				PAIS	Argentina	104	27	27	26	24	Feb 13	n/a	3.56
S016	y	y	y	y	y	n	n	T34	T34	T34	Engineering	India	103	27	24	28	24	Dec 12	n/a	4.44
S017	y	n	y	y	y	n	n	T35; T36	T35; T36	T35	Engineering	Colombia	95	23	24	24	24	Mar 13	Pass	3.56
S018	y	y	y	y	y	n	n	T37	T13		Manufacturing	India	106	27	23	28	28	Apr 13	Pass	4.67
S019	y	n	y	y	y	n	n	T38		T38	Statistics	Italy	106	28	23	30	25	Nov 11	n/a	4.33
S021	y	n	y	y	y	90	n	T37			Manufacturing	Iraq	103	29	22	28	24	Apr 13	With merit	4.22
S030	y	n	y	y	y	200	n				Economics	Peru	98	26	23	24	25	Apr 13	With distinction	3.78
S031	n	n	y	y	y	90	n	T40; T41; T42	T42	T42	Applied Linguistics	China	101	25	24	28	24	May 12	With distinction	3.50
S032	n	n	y	y	y	90	n	T43; T44	T44	T44	Applied Linguistics	China	92	26	24	21	21	Jan 13	Pass	3.89
S033	n	n	y	y	n	90	n	T45			PAIS	Korea	100	24	23	26	27	May 13	Pass	n
S034	y	n	y	y	y	n	n	T48	T32		PAIS	Belgium	112	29	30	24	29	Jan 13	With merit	n
S035[a]	n	n	y	n	n	90	n	T49; T50			PAIS	Japan	102	28	24	28	22	Aug 12	n/a	n
S036	n	y	y	n	y	90	n				Manufacturing	Italy	86	19	22	24	21	Jun 13	With distinction	3.78
S037[a]	n	n	y	n	n	90	n	*T52*			Economics	Japan	107	28	24	25	30	Mar 12	n/a	n
S038[a]	n	n	y	n	n	90	n				PAIS	Japan	109	30	29	22	28	Aug 11	n/a	n
S039	y	n	y	y	y	90	3	T14; *T56*; T58; T59			Manufacturing	Colombia	95	17	22	29	27	Apr 13	with distinction	3.33

Note. y = student did participate. n = student did not participate or attend. EAP tutors are marked in italics.
[a] S035, S037, and S038 were visiting students who had to attend presessional and did not receive a degree.

Suggested citation:

Harsch, C., Ushioda, E., & Ladroue, C. (2017). *Investigating the predictive validity of* TOEFL iBT® *test scores and their use in informing policy in a United Kingdom university setting* (TOEFL iBT Research Report No. 30). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/ets2.12167

Action Editor: Donald Powers

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