ETS TOEFL

Quality Beyond Measure.

TOEFL iBT® Research Report
TOEFL iBT–24
ETS Research Report No. RR–14-44

A Comparative Investigation Into Understandings and Uses of the *TOEFL iBT*® Test, the International English Language Testing Service (Academic) Test, and the Pearson Test of English for Graduate Admissions in the United States and Australia: A Case Study of Two University Contexts

April Ginther

Catherine Elder

December 2014

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, two-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 l 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 (2014-2015) members of the TOEFL COE are:

Sara Weigle - Chair	Georgia State University
Yuko Goto Butler	University of Pennsylvania
Sheila Embleson	York University
Luke Harding	Lancaster University
Eunice Eunhee Jang	University of Toronto
Marianne Nikolov	University of Pécs
Lia Plakans	University of Iowa
James Purpura	Teachers College, Columbia University
John Read	The University of Auckland
Carsten Roever	The University of Melbourne
Diane Schmitt	Nottingham Trent 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

A Comparative Investigation Into Understandings and Uses of the *TOEFL iBT*® Test, the International English Language Testing Service (Academic) Test, and the Pearson Test of English for Graduate Admissions in the United States and Australia: A Case Study of Two University Contexts

April Ginther[1] & Catherine Elder[2]

1 Purdue University
2 University of Melbourne

In line with expanded conceptualizations of validity that encompass the interpretations and uses of test scores in particular policy contexts, this report presents results of a comparative analysis of institutional understandings and uses of 3 international English proficiency tests widely used for tertiary selection—the *TOEFL iBT*® test, the International English Language Testing Service (IELTS; Academic), and the Pearson Test of English (PTE)—at 2 major research universities, 1 in the United States and the other in Australia. Adopting an instrumental case study approach, the study investigated levels of knowledge about and uses of test scores in international graduate student admissions procedures by key stakeholders at Purdue University and the University of Melbourne. Data for the study were gathered via a questionnaire eliciting fixed-choice responses and supplemented with qualitative interview data querying the basis for participants' beliefs, understandings, and practices. The study found that the primary use of language-proficiency test scores, whether *TOEFL*®, IELTS, or PTE, by those involved in the admissions process at both institutions was often limited to determining whether applicants had met the institutional cutoff for admission. Beyond this focused and arguably narrow use, language-proficiency test scores had little impact on admissions decisions, which largely depended on other required elements of applicants' admissions files. In addition, and despite applicants having submitted test scores that met the required cutoffs, survey respondents and interviewees often indicated dissatisfaction with enrolled students' levels of English-language proficiency, both for academic study and for other roles within the university and in subsequent employment. A slight majority at both institutions indicated that they believed the institutional cutoffs represented *adequate* proficiency, while the remainder indicated that they believed the cutoffs represented *minimal* proficiency. The tension created by users' limited use of language-proficiency scores beyond the cut, uncertainty about what cutscores represent, the assumption on the part of many respondents that students should be entering with language skills that allow *success* in graduate studies, and subsequent dissatisfaction with enrolled students' actual language proficiency may contribute to a perception that English-language proficiency test scores are of questionable value; that is, perceived problems reside with the tests, rather than with how test scores are used and interpreted by those involved in the admissions process. At the same time, respondents at both institutions readily acknowledged very limited familiarity with or understanding of the English-language tests that their institutions had approved for admissions. Owing to this lack of familiarity, a substantial majority at both institutions indicated *no preference* for either the TOEFL or the IELTS, counter to our expectation that score users in a North American educational context would prefer the TOEFL, while those in an Australian educational context would prefer the IELTS. The study's findings enhance understandings of test attitudes and test use. Findings may also provide insight for ETS and other language test developers about the context-sensitive strategies that could be needed to encourage test score users to extend their understandings and use of language-proficiency test scores.

Keywords Test score use; English admissions tests; test attitudes; language assessment literacy

doi:10.1002/ets2.12037

The study offers a comparative analysis of institutional understandings and uses of three widely used international English proficiency tests—the *TOEFL iBT*® test, the International English Language Testing Service (IELTS; Academic), and the Pearson Test of English (PTE)—in the academic admissions process at each of two prominent research universities, one in the United States and one in Australia. Each institution exemplifies somewhat different traditions with respect to graduate education, in general, and test use for admission of international graduate students, in particular.

Corresponding author: A. Ginther, E-mail: aginther@purdue.edu

The United States has a longer tradition of enrolling international students than Australia, and the presence of such students at the graduate level was described as early as the 1980s as "vital to certain institutions and to whole fields of advanced study" (Goodwin & Nacht, as cited in Fisher, 1985, p. 64). In Australia, the growth of the international sector in higher education began later than in the United States, following a policy shift in the mid-1980s away from a philosophy of *educational aid* to one of *educational trade*, involving marketing and recruitment strategies geared to boost fee revenue in the face of declining government support for tertiary education (Back, Davis, & Olsen, 1996). The rate of growth in Australia has been extremely rapid, and in the space of only 20 years, Australia has become the fifth largest exporter of tertiary education, with 6.9% of all foreign students worldwide, despite the Australian sector being proportionally much smaller than the sectors of other major export nations, such as the United States (Marginson, 2011, p. 21). Although numbers have stabilized in recent years, the high concentrations of such students in certain programs, including the limited English-language proficiency and limited academic preparedness that some students display both on admission and prior to graduation, pose considerable challenges in both countries (Chow, 2012).

As for the language tests used for admission purposes, the *TOEFL*® test, developed in the 1960s by an independent American nonprofit organization, the Educational Testing Service (ETS), has a much longer tradition of use in the United States, whereas the IELTS, developed in partnership with the British Council, Cambridge ESOL, and IELTS Australia, has far greater currency in Australia. However, the dominance of each test in the respective national context is now being challenged. In Australia, a recent change in government regulations has resulted in the recognition of the TOEFL test and a range of other tests, including a newcomer to the international testing marketplace, the PTE, as alternatives to IELTS for the issuing of international student visas. This change is likely to yield larger numbers of university applicants presenting with scores on TOEFL and PTE than was previously the case. Likewise, in the United States, the inroads made by IELTS (since the early 2000s) and the PTE (more recently) into the admissions testing market have triggered more widespread institutional acceptance of these competitor tests as alternatives to the TOEFL. The fact that institutions in each country may now be presented with a greater diversity of evidence of English proficiency places greater demands on the level of understanding required for effective decision making. The goal of the study is to determine how well equipped institutional score users are to meet these demands.

Such investigation is of particular importance given (a) the increasing numbers of international students undertaking graduate-level study in institutions of higher learning around the world, (b) the importance of English as a vehicle of communication in an increasingly global society, and (c) growing concerns in both Australia and the United States about the limited capacity of many international students to participate effectively in their study programs because of limited English proficiency (Baird, 2010). Increasingly, lack of English proficiency is also being identified as a major obstacle to graduate employment (Birrell, 2006). The study adopts an instrumental case study approach to shed light on users' understandings and practices in relation to the English-language tests used as part of the process of selecting international graduate students at two major research universities. The insights derived from the study are relevant not only within the particular institutions investigated but also in the wider context of English-medium higher education.

The report begins with a literature review, which briefly covers debates around current, use-oriented conceptualizations of test validity and the notion of assessment literacy and what this entails in the university admissions context. The review paves the way for the research questions posed in the current study, namely, the following:

1. How are English-language proficiency test scores used in the graduate admissions process at each institution—particularly with reference to other required elements of a candidate's admissions file?
2. How familiar are institutional test score users with the English proficiency tests (TOEFL, IELTS, and PTE) currently used for admission of international students?
3. What do institutional score users know and believe about language testing and the testing methods used in different language proficiency instruments?

The scene for the study is then set by providing a description of the two sites at which the research was conducted, with particular reference to admissions requirements and procedures for international students in each context. The mixed-methods case study approach used in the study is then described, including an account of the instruments (questionnaires and follow-up interviews) used to gather data and the methods of analysis (quantitative and qualitative) that are used to make sense of participants' responses. Results are then presented and followed by a discussion of the findings in light of previous research. The report finishes with some brief concluding remarks, pointing to the implications of the current

study's findings for both testing agencies and institutions of higher education making use of English test scores in their selection decisions.

Literature Review

Considerable discussion of the meaning, importance, and theoretical status of consequential validity has taken place since the introduction of Messick's (1989) facets of validity framework. Though Messick argues that validity concerns are best addressed under the auspices of the unifying concept of construct validity, the matrix presents facets in which validity is broken down into components associated with test score interpretation and test score use. He argues that validity arguments need to address not only evidence explicating the construct, relevance, and utility of test scores but also the value implications and social consequences that can be associated with test score use. However, the status of consequential validity as a necessary component in the development of validity arguments remains open to debate.

Critics of the inclusion of consequential validity into validity arguments claim that consideration of the consequences of test use unnecessarily complicate discussions of validity, muddling and undermining the scientific basis for the concept (Mehrens, 1997; Popham, 1997). Although Popham and Mehrens acknowledged the importance of consequences, they argue that such considerations are best left outside theoretical discussions of validity and should more appropriately occur within larger public policy domains. Advocates for the systematic inclusion of consequential validity into validity arguments contend that the consideration of consequences is a necessary corollary to actual and/or any reasonably anticipated test score use (Kane, 2002; Linn, 1997; Shepard, 1997).

Bachman and Palmer (2010), in the field of language testing, endorsed the latter position and took it further, proposing that accountability for the consequences of test use should be the primary consideration in developing and using any language test. A similar view of test consequences as the driver for validation efforts is evident in the work of Chapelle, Enright, and Jamieson (2008). These scholars made a detailed case for the proposed interpretation and use of TOEFL iBT test scores and indicated the kinds of evidence that might be used in subsequent investigations of potential or actual use. Thus far, however, given the relatively short history of this new test, there has been little exploration of actual uses and interpretations of the TOEFL iBT scores by different stakeholders.

Educational theorists continue to discuss the relative merits of consequential validity; however, the theoretical issues surrounding what has been referred to as the "Great Validity Debate" (Crocker, 1997) have been eclipsed by the growing acknowledgement that language assessments in educational systems serve social, cultural, and political goals, whether explicit or implicit (McNamara & Roever, 2006), and that test score use and interpretation has serious consequences for students, teachers, school districts, and states. As Kane (2002) observed,

> The traditional view of measurement as an essentially noninteractive monitoring device has been replaced by a recognition that assessment programs can have a major impact on those assessed (Crooks, 1988; Moss, 1998), and more recently, by a conception of tests as the engines of reform and accountability in education. It is the explicit intent of many recent testing programs to promote certain outcomes or consequences, for example to raise standards, to promote changes in curriculum and instruction, and to hold schools accountable (Haertel, 1999). For good or ill, these developments are likely to push policy inferences and assumptions to center stage. (p. 33)

While *tests as the engines of reform* is a phenomenon most commonly associated with mandated K – 12 testing programs like No Child Left Behind in the United States, institutions of higher education are no longer exempt from accountability. However brutal, commonly employed metrics of student success include undergraduate and graduate first-year completion rates and 4- and 5-year graduation rates. With such metrics, selection decisions can be assumed to have a considerable effect on outcomes. Scrutiny of admissions policies and decision-making processes, including how test scores are interpreted and the *assessment literacy* of test users, is therefore warranted.

The term *assessment literacy,* which Stiggins (1991, p. 534) initially coined to refer to the need for mastery of sound principles of assessment by educators in the interests of furthering efforts toward school improvement in the United States, has since been applied to a broader range of contexts and stakeholders (Taylor, 2009). The International Test Commission's (2000) International Guidelines for Test Use propose a set of general competencies required for the users of tests in professional contexts but acknowledges the need for these to be adapted for particular situations. In the field of language testing, the term *language assessment literacy* (LAL; Inbar-Lourie, 2008) has been adopted to denote the particular body of knowledge relevant to users of scores from language tests. However, as Harding and Pill (2013) pointed out, even within

this field, what constitutes an appropriate level of knowledge will vary according to the context of use. In the higher education context, which is the focus of the current study, LAL will involve, at the very least, understanding both the value and limitations of language test scores in the decision-making process, as well as familiarity with instruments used.

The decision to use either the TOEFL test, IELTS,[1] or PTE, or all three tests interchangeably (as is increasingly the case), for selection and/or funding decisions in higher education ideally requires understanding of the test constructs and the meanings of test scores, as well as an awareness of notions of comparability (or lack thereof) between one test and another. Furthermore, setting cutscores on each test is not a simple matter, because language demands may vary widely between individual courses or programs. Good selection practices, as Chalhoub-Deville and Turner (2000) have advocated, entail not only sound decisions about what constitutes an appropriate overall minimum score for university entrance but also some understanding of the subscales and how these might contribute to the selection process for different disciplines. Cutscores should also be subject to ongoing review, via routine monitoring and evaluation of their impact, although it must be said that such evaluations are not straightforward, given the difficulty of isolating language from other factors contributing to language success (see, e.g., Davies & Criper, 1988; Graham, 1987).

In large institutions comprising many different programs, conceptualizations of appropriate language preparation as evidenced by test scores may differ considerably across programs. Furthermore, willingness to articulate those conceptualizations may differ across programs as well, given that examination of test score use in admission procedures involves time, money, and also appropriate preparation and training.

Of the K–12 educational context and the use of standardized scores, Linn (1998) remarked,

> The Test Standards (APA, AERA, NCME, 1985 [1999]) provide guidance and can be used in some instances to hold professionals responsible (for appropriate score use and interpretation), but it would be an unusual legislator or school board member that belongs to one of the three associations that sponsor the Test Standards. (p. 28)

At the university level, outside of psychology and education departments, the same can be said of the many professors and university administrators who are involved in the admissions process. Linn's solution to the problem of responsibility was to argue that the evaluation of the consequences of test score use be partitioned among the stakeholders.

Lane, Parke, and Stone (1998) recommended the use of student, teacher, and administrator surveys to gain information regarding the way stakeholders interpret and use assessment results. They argued that stakeholders' investment in an assessment program will be influenced by their familiarity with test scores, their beliefs about the value of the test scores, and their perceptions of the meaning of the tasks underlying test scores.

Taleporos (1998) found that when teachers, parents, school administrators, and superintendents in New York City were surveyed about what they wanted and needed in a testing program, the interested parties displayed some misunderstandings about the assessment procedures that were in place, along with competing needs and desires about their revision. Information provided by the survey complicated the revision process, but Taleporos argued that it was just as important and meaningful as statistical research data concerning the testing program. Exposing competing interests created an opportunity for open dialog and a more meaningful vetting procedure for all parties involved.

Whereas the value of canvassing score users' understandings and uses of test scores is increasingly acknowledged, explorations of such beliefs and practices in higher education settings are rare. One exception is a small-scale study by O'Loughlin (2011), who employed similar methods to those adopted for the current research to seek information about IELTS test score use and interpretation among particular faculty of an Australian university. O'Loughlin found that although the university provided clear guidelines with respect to selection procedures, the academic and administrative staff displayed considerable range of awareness about those guidelines. More interesting was his discussion of what he termed *folkloric* beliefs on the part of score users about the IELTS, language proficiency, and the contribution of language proficiency to success in university programs. Given O'Loughlin's conclusion that "the decisions made on the basis of applicants' test scores in the particular faculty were poorly informed and were therefore neither valid nor ethical" (p. 159), further exploration of understandings and score uses on a larger scale and across different academic contexts is clearly warranted.

Given the contribution of the TOEFL test in the United States and IELTS in Australia in selection and the recent emergence of the PTE in both countries, monitoring understandings and uses of these tests is critical to ensure (a) that the institutions' interests with respect to graduation rates and/or other metrics of success are optimized, (b) that appropriate policies are established by receiving institutions with respect to the range of abilities represented by the scores awarded to admitted students, and (c) that fairness is guaranteed for individual test takers. This study provides a reasonable starting

point by asking institutional score users what they know about assessment, in general, and how they actually understand and use particular test scores in the process of selecting international graduate students.

Research Questions

The study was designed to address the following research questions:

1. How are English-language proficiency test scores used in the graduate admissions process at each institution—particularly with reference to other required elements of a candidate's admissions file?
2. How familiar are institutional test score users with the English proficiency tests (TOEFL, IELTS, and PTE) currently used for admission of international students?
3. What do institutional score users know and believe about language testing and the testing methods used in different language proficiency instruments?

Research Sites

The research was conducted at two first-tier research universities, Purdue University and the University of Melbourne. As required by our case study approach (Merriam, 1998), considerable background research was conducted at each site to build a picture of similarities and differences between the two universities in their policies and practices in relation to international graduate students. The advantage of the case study approach is that it enables the researcher to ground observations within a particular context and to understand beliefs and decisions in relation to local social networks, policies, and historical conditions (Merriam, 1998; Yin, 2008). Nevertheless, the approach adopted here should be considered instrumental, rather than intrinsic. According to Stake (2000), an intrinsic case study focuses exclusively on the particular case, whereas an instrumental approach is designed for the purpose of illuminating a larger issue or set of issues extending beyond the context of concern. It is our intention not simply to describe beliefs and practices within the selected institutions but also to draw insights from the particular contexts, which might be useful in understanding how language test scores are perceived and used within the broader context of educational and administrative decision making in higher education.

To gain understanding of the two sites, we drew on a variety of available sources, including minutes of relevant committees and published reports and documentation provided by relevant departments, such as the Finance and Planning Office at the University of Melbourne and Purdue's Office of International Students and Scholars (Purdue University, 2011). A brief digest of the information gathered is provided in the following to assist the reader in understanding the two contexts and making sense of subsequent findings. Further details, including the history of English-language testing in the admission process at each institution, can be obtained from the researchers.

International Student Enrollments

Purdue is a large land grant university[2] in the United States with an overall current enrollment of around 40,000 students. Melbourne is the second oldest public university in Australia, with an enrollment of around 50,000. Both universities consistently rank among the top research universities for international graduate student enrollment, with Purdue enrolling 2,200 such applicants in 2011 and Melbourne enrolling just under 5,000 in the same year. International graduate applications at Melbourne were approximately double admissions, indicating a rather liberal overall admissions rate (lower selectivity) than at Purdue (53% vs. 23%). This may be due, on one hand, to the larger number of students applying to Purdue than to Melbourne and, on the other, to the higher ratio of international students enrolled in fee-paying master's degree coursework rather than in master's and PhD programs by research (approximately 4:1).[3] Such courses are more prevalent at Melbourne than at Purdue and generally have less stringent entry requirements than doctoral and master's research degree programs.

Minimum English-Language Requirements

Both institutions now accept the TOEFL iBT test, IELTS, and PTE for English screening purposes, although the TOEFL test has a stronger tradition of use at Purdue (as at other U.S. universities), and the same is true for IELTS at Melbourne (and elsewhere in Australia). At Purdue, the current minimum English entrance requirements for international graduate

students on the three approved tests are as follows: TOEFL iBT, 77+ writing 18, speaking 18, reading 19, listening 14; IELTS, 6.5 with no subsection minimums; PTE, 58 with no subsection minimums. The subsection cutoffs on the TOEFL test were set at a 2-day in-house standard-setting session conducted by Purdue's Graduate School in 2005, with subsequent adjustments to bring the minimum total score into line with that of other comparable institutions. No standard-setting exercises were conducted for the IELTS or the recently accepted PTE. Cutoffs on these tests were set to align with the other institutions that compose the Big 10.

At Melbourne, entry requirements have been modified over the years in response to local departmental submissions to the university's selection committee. No standard-setting study has been conducted, however, either for IELTS or for any of the other tests. As at Purdue, cutoffs are set with an eye to practices at other comparable universities. The current university-stipulated minimums for graduate study are as follows: TOEFL iBT, 79+ writing 21, speaking 18, reading 13, listening 13; IELTS 6.5+ writing, speaking, reading, and listening 6; PTE, 58+ no communicative skill below 50. The requirements, however, vary across programs, with PhD and master's programs, including philosophy, creative writing and publishing, teaching, dentistry, and nursing and medicine, setting their cutscores higher. The highest threshold is a TOEFL score of 94 (+W27, S18, R13, and L13), an IELTS of 7 (+W7, S6, R6, and L6), or the equivalent on other recognized tests. Some faculties, such as engineering, allow lower entrance scores for postgraduate study, subject to completion of a bridging program prior to entry or a prescribed English study program in the early stages of the degree coursework. Full details are available online (University of Melbourne, 2014).

It can be seen that minimum cutscores are roughly comparable across sites. Although Purdue requires considerably higher minimum scores for TOEFL reading than Melbourne, the total cutscore is lower. Total IELTS and PTE cutoffs are the same at each site, the only difference being that Melbourne stipulates cutoffs for each subsection, whereas this is not the case at Purdue.

Admissions Procedures

All graduate applicants for Purdue are required to submit their materials to Purdue's Graduate School, which serves merely as a clearinghouse for applications. The Graduate School tracks applications for statistical reporting but plays no part in making actual admissions decisions. Completed files are forwarded to the department or program in which the applicant seeks admission. Decisions about graduate admissions are made by departmental or program committees in a decentralized process. Whereas some programs join together to form a central committee, for example, 10 programs in the life sciences have formed a central graduate admissions committee, more commonly, each department (e.g., mechanical engineering, electrical and computer engineering, and materials engineering) has its own admissions committees within a college (e.g., the College of Engineering). Within single departments, it is often the case that individual programs have independent committees. For example, in the English department, creative writing, English language and linguistics, literature, rhetoric and composition, and second language studies each has its own independent admissions committee. However, all offers of funding must be approved by the director of first-year composition, and all incoming students are required to teach Introductory Composition for at least 1 year.

At Melbourne, the process of monitoring and guaranteeing minimum English entry levels is conducted by administrative staff in the admissions office, who also check the applicant's file for academic eligibility. The files are then forwarded to the relevant faculty or graduate school for further consideration. As at Purdue, the mechanisms for selection vary considerably across the university, with some faculties convening centralized committees that operate across programs and others organizing these matters locally at the particular program level. For master's by coursework degrees, admissions decisions are generally made by a faculty committee in consultation with the relevant course convenor. In the case of doctoral students, admissions decisions are made on a student-by-student basis by laboratory heads (in disciplines like science) or by thesis supervisors with the endorsement of the head of school or relevant disciplinary unit. Note, however, that, differently from Purdue, all the student files are checked centrally before any offer is made, and a special case has to be made to the university's admissions office for any waiver of stipulated minimum requirements, whether related to English proficiency or the candidate's academic status.

Conditions of Funding

An important difference between the two institutions relating to work opportunities for graduate students is worth noting at this point because it may influence views that are canvassed in this study regarding the importance of English in the

admissions process. Whereas at Purdue, and in U.S. universities more generally, many graduate students are funded with a stipend on the expectation that they will participate in part-time teaching assistant duties for undergraduate programs, this is not normally the case at Melbourne. Here, and indeed at all higher education institutions in Australia, graduate students are eligible for a range of government- or university-funded scholarships that offer fee remission, health insurance, and, in some cases, a tax-free annual stipend for the duration of the degree program. Though the conditions of most scholarships allow students to work for up to 20 hours per week throughout their studies, students do not routinely carry out tutoring work for the university.

Postentry English-Language Support

Given the key role of graduate students in undergraduate instruction at Purdue, English support comes in the form of a testing and instructional program in English for prospective international teaching assistants, which is required for those who score below a given threshold on the locally developed Oral English Proficiency Test (Purdue University, 2012). A pass on this test or in the language support class is a prerequisite for employment as a teaching assistant. Several classes are offered in association with specific programs or departments, such as in electrical and computer engineering, which gives a writing proficiency screening to both L1 and L2 speakers and offers support for those (mainly L2 learners) who do not pass. English-language support at Melbourne is available in various parts of the university for those who score below a certain threshold on the locally developed Diagnostic English Language Assessment. Available courses range from a credit-bearing generic English for academic purposes (EAP) program to discipline-specific tutorials embedded within students' study programs to one-on-one assistance with academic writing. Participation in these support options is, however, limited and inconsistent, perhaps because the programs are generally not compulsory for graduate students.

Method

The study adopted a sequential exploratory mixed-methods design (Creswell & Clark, 2011), involving three phases of data collection, namely, (a) a literature review and a document analysis to provide information about the concepts relevant to the study and the research sites investigated (see the previous section), (b) a survey of faculty involved in the admissions process, and (c) postsurvey interviews with selected faculty informants. The methods of analysis included both descriptive statistics (for the survey data) and thematic analysis (for the postsurvey interviews). The mixed-methods design was chosen to increase the interpretability and meaningfulness of the enquiry by seeking "elaboration, enhancement, illustration, clarification of the results from one method with the results from the other method" (Greene, Caracelli, & Graham, 1989, p. 259).

In the following sections, we describe in detail the process of designing and administering the survey. We then outline our approach to designing and conducting the interviews and to analyzing the interview data.

The Survey and Survey Development

The survey development consisted of three phases: designing of questions (informed by presurvey interviews), pretesting, and piloting. Each of these phases is described subsequently.

Survey Development I: Designing the Questions

The questionnaire for this study was designed to elicit information about a range of issues deemed to be relevant to uses and interpretations of English test scores by faculty dealing with the selection of international graduate students within the two universities under investigation. The content was informed not only by a preliminary literature review of the role of English in academic study and the use of English tests in the admissions process but also by 26 presurvey interviews. The latter were conducted to better understand users' understandings in the graduate admissions process at both universities, thereby enhancing the content validity of the various items before we fully committed to their use.

A description of the survey content is provided in the following subsections. A full copy of the survey is available on request from the researchers. The first four sections of the survey were designed to gather background information about the participants at each site. The remaining sections elicited information of direct relevance to the research questions about score users' understandings, attitudes, and practices in relation to English tests in the university admissions context.

Participant Background

This section of the questionnaire canvassed data about participants' roles in the university, home colleges, departments, and programs and whether they have specific responsibilities for graduate student admissions. Such distinctions were important because they could reveal variations in understandings and attitudes according to role and, accordingly, how and for whom information about English tests might need to be packaged. The section contained 11 questions.

Characteristics of Participant's Program

This section elicited information about the degrees granted by respondents' programs and about the international–domestic mix of those applying to them. We recognized that many participants are involved in more than one degree-granting program (e.g., coursework master's, research master's, or PhD), and, as stated earlier, we asked respondents to identify their professional programs. To reduce the time involved, participants were instructed to respond in relation to the program with which they have more experience or about which they know the most. Questions canvassed information about funding arrangements for international students and about English-language proficiency requirements. Though we are aware that North American and Australian universities differ markedly from one another in funding arrangements for graduate students, we wanted to ascertain that this is indeed a uniform situation across programs. This section contained six questions.

Role of English in First-Year Academic Success

This section elicited information about respondents' beliefs about the importance of English-language proficiency for international graduate students in their first year of study by way of background for the research. In addition, three questions addressed the respondents' perceptions of the language proficiency admissions requirements and the levels of proficiency that required admissions scores might represent. After providing information about language requirements at the institution in question, a question was also asked about what this language proficiency admissions requirement represents, given that in Banerjee's (2003) and O'Loughlin's (2011) studies, and in our preliminary interviews, some of the participants seemed unaware that the required scores were intended as minimum standards rather than an indication that a candidate's English was satisfactory for all aspects of academic study. The final question asked about respondents' beliefs about the rationale for setting cutscores. This section contained five questions.

Available Funding for International Graduate Students

This section asked respondents to identify the sources of funding available to international graduate students in their programs (fellowships, teaching assistantships, research assistantships) and whether students are admitted without funding. This section contained four questions.

Making Admissions and Funding Decisions for International Graduate Students

This section addressed language proficiency tests and their role in selection. It queried participants' views about the importance of English test scores (both overall scores and subsection scores), along with other aspects of graduate attributes or experience (e.g., *Graduate Record Examinations*®; *GRE*® score, academic transcript, work experience) in making funding decisions, including for scholarships and research or teaching assistantships. Respondents who offered different kinds of student funding were invited to respond to the questions for each case. Responses are in four categories, with a *not required* option supplied for components of the candidate's file or background that did not figure in selection for particular programs. This section contained seven questions, but three of them had multiple subcomponents. For example, when respondents were asked about English-language proficiency test scores, they were asked to rank the importance of the total score along with the four traditional subscale scores—reading, writing, speaking, and listening—when making admissions decisions. When asked about the GRE tests, they were asked to rank the importance of the verbal, quantitative, and analytical scores. When asked about candidates' admissions files, they were asked to rank the importance of 10 separate components.

Departmental and Program Standards for Admissions and Funding

This section asked respondents to indicate whether their departments or programs use standards (higher, lower, or the same) for admissions and funding that differ from the standards of the institution. We also asked whether any of the respondents had sought a waiver of the institutional language-proficiency requirements. This section contained only two questions, but the first contained five subquestions.

English-Language Testing Methods

This section focused exclusively on language assessment and canvassed respondents' views regarding methods for assessing English. The methods covered are those featured in the three main tests currently used for selection, namely, IELTS, TOEFL iBT, and PTE, although the specific tests were not named at this point. Our aim was to establish respondents' views about such controversial issues as computer scoring, independently of any reactions they may have had to particular *brands*. In light of feedback from the presurvey interviews, any technical jargon used to describe different modes of assessment (e.g., semidirect) was replaced with lay wording. Dichotomous response options (*favorable* or *unfavorable*) were provided for this question, with an additional option for those with no opinion on these matters. This section contained nine items.

Views and Preferences for Current Tests

Respondents' familiarity with the main English tests used for academic selection was explored in this section, with four Likert-type scale response options provided, ranging from 1 (*very familiar*) to 4 (*not familiar*). For those respondents who stated a preference, we then asked them to indicate, by selecting from a list of options, to what their preference for a particular test might be attributed. The list of reasons included in association with this item is based on those provided by respondents at the presurvey interview stage. We asked those respondents who selected *no preference* to write in their reasons. This section contained four questions, but the question asking for a reason for preference contained 19 subquestions.

General Assessment Literacy and Interest

In this final section of the questionnaire, respondents were asked about their level of knowledge and interest with regard to language testing and assessment. Although it is widely believed that lay users of language test scores have folkloric beliefs (O'Loughlin, 2011) and should become more assessment literate (Chalhoub-Deville & Turner, 2000), there is little evidence of what they know and whether they perceive such knowledge to be important or relevant. There is a further question asking what kinds of information (if any) respondents might like to receive about any of the three admissions tests named in the questionnaire. The items included in this latter question are based on information generally provided in score user manuals. An *other* option is included to capture any additional information needs. Responses to this question may offer useful information for test developers about what informational needs to prioritize. This section contained four items, and the third item contained six subquestions. The fourth and final item asked respondents whether they would be willing to participate in a follow-up interview.

Survey Development II: Pretesting

The final draft survey was pretested with a small sample of respondents with different roles and from different disciplines within each university. This process led to the inclusion of additional response options, changes in the wording of some questions, and exclusion of some questions that were found to be unclear or unworkable. The ordering of the questions was also changed so that questions asking for participants' familiarity with and interest in language tests and language testing were left to last. This was to avoid a situation in which any sense of discomfort regarding lack of such interest or knowledge might result in participants opting out of the questionnaire at an early stage.

A further decision made as a result of the pretest experience was to design slightly different, but nevertheless comparable, versions of the questionnaires for each institution, given differences in admission procedures and terminology. For example, the term *advisor* used in the American system was substituted with *supervisor* in accordance with Australian terminology. Items also differed when the organization of the universities required that we use different language; for

example, the first question on the Purdue survey asks about the respondents' *college*, whereas the first question on the Melbourne survey asks about the respondents' *school*. With respect to graduate admissions, the questions on the Purdue survey asked about service on local departmental or program-level graduate admissions committees; the questions on the Melbourne survey asked about involvement with the admissions process more generally, given that this process varies widely across the university and does not always include a local committee structure. Furthermore, we asked an additional question of Purdue respondents: Item 13 asked them to provide the name of the professional program if they were primarily associated with a professional certificate program rather than a degree-granting program. In all other respects, the instruments were identical. The final draft survey consisted of 52 items for Purdue and 51 items for Melbourne.

Survey Development III: Piloting

A third version, matching the version designed for Purdue, was created for pilot administration at East Carolina University (ECU). Piloting was conducted at ECU from February to March 2012. The ECU pilot was administered twice. The draft survey was sent to selected faculty members who were identified as having some interest in testing issues and as able and willing to contribute to the pilot effort. Respondents commented on several issues that were addressed with revisions. First, they commented on the length and repetitiveness of the instrument, particularly with respect to the questions concerning required language proficiency scores for admissions, for awarding assistantships, and for awarding fellowships. An option allowing respondents to indicate that the criteria were the same was added. Second, for some items, respondents were uncertain whether the question was asking about institutional policy or personal opinion. The questions where this issue might be a problem were clarified. In addition, several commented that they wanted a *Don't know* option for all items, and this option was provided where applicable. Third, several indicated that they were most familiar with the pencil-and-paper (PBT) and computer-based (CBT) versions of the TOEFL test and less familiar with the Internet version (TOEFL iBT), but the iBT was the only TOEFL test that was included. Along these lines, several also indicated that they were more familiar with the score ranges associated with particular versions of the TOEFL test, rather than with the names of the tests themselves, and were uncertain about which scale belonged to which test. At this point, we decided to include the three versions of the TOEFL test along with the total score scale ranges in parentheses to aid in respondents' selection of the test(s) with which they were familiar or unfamiliar. After these revisions were made, the revised instrument was distributed to all graduate faculties excluding those who had responded in the first round through ECU's Qualtrics site; therefore, the final draft instrument matched the final Purdue version in both content and administration. Seventy-one responses (a 10% response rate) were received. No additional problems were identified.

Survey Administration

The two parallel sets of questionnaire items (for Purdue and Melbourne, respectively) were revised within the Qualtrics program at Purdue University's Qualtrics site. Survey administration at both sites, Purdue and Melbourne, was conducted through the Purdue Qualtrics site.

Data collection took place between January and June 2012. All graduate faculties were invited to respond,[4] and prospective survey participants were contacted via e-mail. As neither institution was willing to share the graduate faculty e-mail list, these were created by visiting departmental websites and then compiling e-mail lists from all positions listed in each department at both institutions. This was not a perfect solution, given that departments list both faculty and administrative positions, and it is often impossible to determine whether the staff member listed holds a faculty or administrative position or which faculty positions are involved in teaching exclusively in graduate rather than undergraduate degree programs. A further problem is that, in some departments, the director of graduate studies might be held by a faculty member, but, in other departments, a similar position might be held by an administrative staff member (i.e., a nonfaculty position). An additional complication is that many faculty positions at both institutions are clinical or research positions for which there are no actual teaching responsibilities. In the end, invitation letters were sent to many faculty members, those holding clinical, research, and undergraduate teaching appointments, who would actually never have been involved in the graduate admissions process.

Those interested in participating in the study and taking the survey clicked on a link indicating agreement to participate in the study. This link took them directly to the survey. Those who wished to opt out of the study could either click on a link that removed them from the list of participants or simply ignore all e-mails pertaining to the study.

Participation

At Purdue, the invitation to participate was sent to 2,678 faculty and staff, and 232 respondents completed the survey—a response rate of 8.6%. However, if calculated on the actual number of graduate teaching faculty of 1,687, the response rate is 13.7%. At Melbourne, the invitation to participate was sent to 2,337 faculty and staff, and 249 completed the survey—a response rate of 10.7%. If calculated on the actual number of graduate teaching faculty, the response rate is, again, likely to be higher. However, no official figures are available on the number of faculty members teaching exclusively in graduate programs.

At Purdue, the majority of the respondents were members of the College of Liberal Arts (22%). The Colleges of Agriculture (16%), Engineering (14%), and Science (13%) followed. Each of the remaining colleges was represented by less than 10% of the participant pool. At Melbourne, the majority of respondents represented the Faculty of Medicine, Dentistry, and Health Sciences (20%). The next largest group selected Other (14%), then Science (11%), as their faculty. Each of the remaining faculties was represented by less than 10% of the participant pool. Within colleges at Purdue and faculties at Melbourne, there was a wide range of representation of departments or schools and programs. The majority of respondents at Purdue (85%) indicated that English was their first language, and the percentage of English L1 speakers at Melbourne (87%) was comparable. Around one third of respondents at both Purdue and Melbourne indicated proficiency in at least one second language, with English being the most commonly nominated second language at both institutions.

It should be noted that, at both institutions, a substantial number of invitation recipients started but failed to complete the survey: at Purdue, of the 278 who started, 232 completed (−46), and, at Melbourne, of the 365 who started, 249 completed (−116). At both institutions, we received e-mails from graduate faculty who reported that they dropped out before completing for a range of reasons, including lack of involvement with graduate student admissions decisions and, in particular, a lack of familiarity with language tests. When appropriate, we encouraged these faculty members to complete the survey despite their limited familiarity with language proficiency instruments, given that we were conducting the survey because their familiarity (or lack of familiarity) with the instruments involved was an issue that the survey addressed. It is impossible to determine how many prospective respondents actually did return to the survey after being encouraged to do so. It seems likely, given the preceding considerations, that the survey sample is not fully representative of the wider population of graduate faculty at each university and that there is a bias in favor of those who are both more experienced with the student admissions and either more concerned about language proficiency issues or more knowledgeable about language tests than their colleagues. There may also have been a higher response rate among native English-speaking faculty or those with strong feelings about the role of English in student admissions and in the academic experience more generally. This should be borne in mind in evaluating the study's results.

Survey Analysis

Survey data were analyzed descriptively using reports generated by the Qualtrics program.

Postsurvey Interviews

The follow-up interviews were semistructured in format, centering around the individual's responses to the topics covered in the survey, which had been summarized on background sheets prepared for the interviewer, with an asterisk beside any response that ran counter to the general trend in the survey responses or appeared of interest in some other respect and might be worthy of further probing.

Interview Administration

Within 2 months of their completion of the survey, a sample of the survey participants who had indicated their willingness to participate in a follow-up interview were contacted by e-mail and invited to schedule an initial interview. This e-mail also contained information on the voluntary nature of the study and the benefits and risks of participating. In drawing the sample, an attempt was made to include faculty members from as broad a range of disciplines as possible. All who were invited agreed to participate. All procedures required by both universities to ensure the confidentiality of participants' responses were followed.

Interviews were conducted jointly by one of the chief investigators and a research assistant in the first instance and, subsequently, by the research assistant alone. Interviewees were first asked in more detail about their role within their department or program and to sketch the processes in place for selection of international graduate students. They were then asked to provide further detail about their survey responses to particular questions, for example, why they felt that English was critical for success in the first year of the program and at what point (if at all) in the admissions process they looked at English proficiency test scores and/or any other evidence in the applicant's file relevant to making decisions about the applicant's ability to meet the language demands of the program in question. Participants were encouraged to elaborate their responses with reference to particular instances, for example, cases where minimum English entrance requirements had been waived or raised. We also asked further questions about participants' knowledge of language testing, as well as their experience with and opinions of particular tests. At the end of each interview, respondents were given the opportunity to make further comments on any aspect of the university's English-language policies that they considered important.

Thirty interview sessions, 16 at Purdue and 14 at Melbourne, were held. Each lasted between 30 minutes and 1 hour. Fourteen academic staff participated individually in the interviews at Melbourne. Two of the 16 sessions at Purdue were conducted with two participants simultaneously.

Analysis of Interview Data

All interviews were audio-recorded and subsequently transcribed and segmented by turn, yielding a total of 2,700 turns. The decision to adopt the turn as the segment for subsequent coding followed a review of various approaches to segmentation and consideration of the purpose of the analysis, which was to identify broad content categories in participant responses, rather than, for example, to quantify particular linguistic features of respondents' speech. Scrutiny of the data revealed that, for the most part, participants dealt with a single topic or content category in response to the researcher's prompt or question and only moved to a subsequent topic in a separate turn. Although there were occasions when participants shifted topic within a turn, it appeared to be difficult to tease the different topics or content categories apart without losing the context necessary to interpret what was being said (see Brice, 2005, for further discussion of this issue). Thus, a decision was made to allow more than one code to be applied to segments, as required.

A tentative thematic coding scheme was developed by the chief investigators following a preliminary scan of the transcripts to identify salient topics or content categories relevant to the study's research questions. This coding scheme was subsequently applied to the data by a research assistant with certain refinements introduced to deal with subthemes not accounted for in the original system.

The final thematic categories (and subthemes) in the coding system are as follows:

- Admissions process and the use of test scores (Research Question 1)
    - Awareness of or views about cutscores or standards for entry
    - The place of English proficiency in the process of selecting students
- Language tests or language testing (Research Questions 2 and 3)
    - Beliefs about and attitudes toward particular language tests (TOEFL/IELTS/PTE)
    - Beliefs about and attitudes toward language tests in general
    - Beliefs about and attitudes toward particular modes of testing
- Other

Note that the category titled Other was used for researcher questions and comments and also for any interviewee responses that were deemed peripheral to the major themes of the study.

Three interview transcripts were randomly selected and double coded by a research assistant. Intercoder agreement for these transcripts averaged 94% for the first coding and dropped to 83% when both the first and second codings were counted. This was because there was not always agreement on the need for double coding; that is, even though the two coders agreed on a salient feature for a segment, if one double coded the segment while the other did not, it was counted as a disagreement. Any instances of disagreement were checked and were able to be resolved by one of the chief investigators in almost every case. Although the number of interviews that were double coded constituted only 10% of the data set, the agreement rate was considered sufficiently high to obviate the need for further reliability checks.

Results and Discussion

Survey Data

The following sections present summaries and a discussion of what we found most interesting and/or noteworthy in the survey results. Often response categories were collapsed, for example, strongly agree and agree may be combined, when it made sense to do so. Within each section, we refer to tables (see the appendix) where the results are also presented. All of the results are reported as percentages, with the number of respondents in parentheses. The numbers of no-response values are presented for each question, and reported percentages are calculated for the number of respondents by item. We had many more no-response values for respondents at Melbourne than at Purdue. This occurred in part because of a failure to include a *Don't know* option in some of the Melbourne items, owing to an editing error. Where this occurred, the absence of the *Don't know* option is marked with a dash in the table.

Final reports for Purdue and Melbourne generated by Qualtrics, along with all the written responses, are not included in this report but are available on request.

Respondents' Service on Graduate Admissions Committees

A slightly larger percentage of the respondents at Purdue than at Melbourne reported that they had served as an advisor to an international graduate student (78% vs. 69%), and, again, a slightly larger percentage of the respondents at Purdue reported that they were currently advising international graduate students (58% vs. 54%). Service on a graduate admission committee (past or current) was more common at Purdue than at Melbourne, most likely because of the more local organization of these committees at the former institution. At Purdue, each program typically organizes its own committee, whereas, at Melbourne, as noted earlier, there is an initial university-wide admissions screening process for international students, at which eligibility issues are dealt with and selection recommendations are made. There is also considerable variation across programs in both the amount of local involvement in admissions decision making and in the particular procedures followed. All responses to questions are reported in Table A1.

Fifty-three percent of the respondents at Purdue and 42% of the respondents at Melbourne wrote in the number of international graduate students they were advising or supervising. At both institutions, the majority (85%) of the respondents who responded to this question indicated that they were advising 1–5 international students, with the remainder (9%) indicating they were supervising 6–9 students or 10 or more students (6%). The final question in this section asked whether respondents had ever discussed the meaning of English-language proficiency test scores with colleagues. At both institutions, more than half of the respondents responded affirmatively (68% at Purdue and 51% at Melbourne). Graduate advising loads are reported in Table A2.

Characteristics of Respondents' Programs

A discussion follows.

Which of the Following Degrees Does Your Program Grant?

Respondents at both institutions indicated that their programs offered both coursework master's degrees and professional degree options. In fact, the coursework option was common at both institutions, with 77% of the respondents at Purdue and 60% of the respondents at Melbourne reporting the existence of coursework master's and professional degrees within their programs (Table A3). This was somewhat at odds with our the claim (see Research Sites) that coursework master's degrees are more common at Melbourne and may be a function of the particular sample of survey respondents. Conversely, when asked to indicate the program with which they were most familiar, respondents at both institutions indicated that they were most familiar with research degrees—those that require a major research thesis for an MA or PhD (84% at Purdue and 60% at Melbourne; Table A4). Only a relatively small percentage of respondents at Purdue (14% or 33%) indicated greater familiarity with coursework degrees; at Melbourne, the percentage was higher (34% or 85; Table A4). We broke down responses on this question by degree type (research vs. coursework only) but were unable to detect any differences in response patterns with respect to admission standards. Therefore, all subsequent responses are examined based on the entire respondent pool.

International Graduate Applications and Admissions

According to participants' self-reports, a slightly larger percentage of international applicants are likely to be admitted into graduate programs at Purdue (Table A5). The percentage of international students currently enrolled across institutions was similar. At Purdue, 9% of the respondents and, at Melbourne, 12% reported international enrollments of less than 10%; at Purdue, 40% and, at Melbourne, 34% of the respondents reported current enrollments ranging from 10% to 50%; at Purdue, 27% and, at Melbourne, 19% of the respondents reported program enrollments of greater than 50% (Table A6).

English-Language Proficiency and Academic Success

A discussion follows.

Role of English in First-Year Academic Success

There was considerable agreement across institutions regarding the importance of English-language proficiency with respect to academic success. When asked whether content knowledge was dependent on English-language proficiency, 69% of the respondents at Purdue either agreed or strongly agreed, whereas 62% of the respondents at Melbourne either agreed or strongly agreed (Table A7). When asked whether they believed that English-language proficiency was critical for students to perform adequately in the first year of their programs, agreement or strong agreement was signaled by the vast majority of respondents: 79% at Purdue and 77% at Melbourne (Table A8).

Perceptions of Language Proficiency Requirements

In response to the question asking their perceptions of what the English-language admissions requirements represent, 52% of the respondents at Purdue indicated that they understood the language proficiency requirements set by the university as *minimal* English-language proficiency requirements, while 38% of the respondents indicated that they believed the requirements represent *adequate* English-language proficiency. Although the terms were not defined and may have been ascribed different meanings by the respondents, it was surprising that 3% of the respondents indicated that they believed the English-language requirements represent *advanced* English-language proficiency. Eight percent of the respondents did not answer, perhaps indicating that they did not know what the requirements represent.

At Melbourne, respondents were less likely to indicate that they understood the English-language requirements as *minimal* with respect to English-language proficiency (30%). Forty-two percent indicated that they believed that the requirements set by the university represented *adequate* English-language proficiency, while 9% indicated that they believed the requirements represent *advanced* English-language proficiency. Nineteen percent failed to answer, perhaps again indicating that they did not know. These responses are reported for both institutions in Table A9.

Graduate School English-Language Proficiency Requirements for Admission

Responses across institutions were comparable in response to the question asking whether they considered the requirements *too low, appropriate,* or *too high.* Just under one third of respondents at both Purdue and Melbourne indicated that they believed the requirements to be *too low*; just over one third at each institution believed the requirements to be *appropriate*; less than 1% at Purdue and only 2% at Melbourne indicated that they believed the requirements to be *too high.* At Purdue, 22% and, at Melbourne, 11% indicated that they did not know. Despite our having included a *Don't know* option for this item, 8% at Purdue and 19% at Melbourne did not answer (Table A10).

When asked to speculate about the motivations on the part of the university with respect to English-language proficiency requirements, respondents were asked to indicate whether they agreed or disagreed with seven possible motivations. When ranked in terms of the largest percentage of respondents either strongly agreeing or agreeing, *A desire to ensure adequate language proficiency of all admitted students* was accorded the highest ranking at both institutions. In light of respondents' tendency to identify language proficiency requirements as too low, respondents' willingness to agree with the idea that the university's intention is *To ensure adequate language proficiency* seems a bit odd, at least at first blush. It may be the case that respondents appreciate the intention but are also aware of problems with its realization. *A desire to align with peer institutions, A desire to secure a large pool of international applicants,* and *A desire to ensure a high rate of*

degree completion were also selected as likely motivations at both institutions, attracting between 52% and 69% agreement (although degree completion was ranked higher at Melbourne than at Purdue). Fewer respondents at both institutions agreed with *A desire to increase revenue, A desire to select only the best graduate students,* and *A desire to increase class size,* yet these remain viable options for at least 30% of the respondents. These responses are presented in Tables A11 and A12.

Funding for International Graduate Students Across Institutions

Respondents differed across the universities across the series of questions asking how international graduate students were funded. At Purdue, as expected, teaching assistantships were reported as the most common form of support (75%); however, research assistantships and fellowships or scholarships were identified as available options by at least 60% of the respondents (Table A13). At Melbourne, fellowships or scholarships were reported as the most common form of funding (60%), while research assistantships (17%) and teaching assistantships (15%) were identified by many fewer respondents (Table A14). This is in keeping with what was reported earlier, namely, that employment opportunities for graduate students as tutors or research assistants are by no means the norm at Australian universities. The universities also differed in terms of admissions for unfunded students: Purdue respondents were more likely to report that grad students were admitted without funding (75%), as opposed to Melbourne, where only 48% of the respondents reported that students were admitted without funding (Table A15).

Making Admissions and Funding Decisions for International Graduate Students

A discussion follows.

English-Language Proficiency

When asked to consider the importance of English-language proficiency test total scores and subscale scores with respect to admissions decisions, the vast majority of the respondents at both universities ranked each of the components of reported scores as either *very important* or *important.* At Purdue, the percentage of respondents who ranked the components of language proficiency scores as *very important* or *important* ranged from a high of 86% for a total score to 79% for a listening subsection score. At Melbourne, the percentage of respondents who ranked the components of language-proficiency scores as *very important* or *important* ranged from a high of 92% for a reading subsection score to 87% for a speaking subsection score. The percentages of respondents ranking particular subscales differed somewhat across institutions; that is, a slightly larger percentage of the respondents at Purdue attributed importance to a total score, whereas a very slightly larger percentage attributed importance to a writing subscale score at Melbourne. However, at both institutions, English-language proficiency was considered either a *very important* or *important* consideration with respect to admissions decisions. Finally, overall, a slightly larger percentage of respondents attributed importance to English-language proficiency scores at Melbourne (Tables A16 and A17).

The GRE

When asked to consider the importance of the GRE with respect to graduate admissions decisions, the percentage of the respondents who reported GRE scores as either *very important* or *important* fell slightly as compared to the percentage of respondents who ranked English-language proficiency in a similar manner. At Purdue, the percentages of respondents reporting GRE scores as either *very important* or *important* ranged from a high of 74% for the GRE Quantitative Reasoning score to 69% for GRE Analytical Writing (Table A18). Contrary to popular belief, at Purdue, the GRE is not required by the Graduate School for admissions purposes; nevertheless, the GRE remains a requirement for approximately two thirds of the departments and programs that have graduate programs (T. Atkinson, personal communication, 2012). At Melbourne, the percentages of respondents reporting GRE scores as either *very important* or *important* ranged from a high of 67% for the GRE total score to 62% for the GRE Quantitative Reasoning score (Table A19). The percentages of respondents reporting GRE scores as important for admissions decisions at Melbourne was unexpected, given that this examination, although listed as a requirement for certain degrees offered by the Graduate School of Business, is not generally required for those applying to study in other disciplinary areas, and, indeed, this test is not widely known in the Australian context. This result may indicate a response bias that inflates the importance of all tests with respect to admissions decisions.

Candidate Admissions File

When asked to consider other aspects of a candidate's admissions file, the largest percentages of respondents at both institutions indicated that a candidate's transcript or academic record was either *very important* or *important*. However, there are some interesting differences in terms of the importance associated with different components of a candidate's file across institutions. At Purdue, the statement of purpose was considered either *very important* or *important* by 84% of the respondents, whereas, at Melbourne, only 65% of the respondents considered the statement of purpose similarly, presumably because such statements are not a required component of the graduate application package. In response to "Any form of oral interview," only 50% of the respondents at Purdue considered this component *very important/important*, while 77% of the respondents at Melbourne attributed importance to this component of the candidate's file, perhaps reflecting the sample bias toward the School of Health Sciences, where oral interviews are a key component of the selection process. At Purdue, some departments reported conducting telephone or Skype interviews, especially when applicants were being considered for a teaching assistantship, but apparently this practice has recently become policy as a final check in the admission process for all international PhD applicants at Melbourne. The difference in respondents reporting letters of recommendation as *very important/important* across institutions was also interesting: 79% at Purdue and only 61% at Melbourne. Differences in the percentages identifying the writing sample or research proposal as *very important/important* also existed across institutions: 59% at Purdue but 73% at Melbourne. This is possibly because, as noted under Research Sites, many PhD programs at Melbourne do not include a major coursework component during which the student can develop his or her research ideas under supervision, making the preparation of a draft research proposal from the outset a more critical element in the selection process. Another difference can be seen with respect to awards from previous institutions, as these were not given much credence by evaluators at Purdue (37% *very important/important*) as compared to 65% at Melbourne. At both institutions, teaching experience was not considered *very important/important* by many of the respondents (only 22% at Purdue and 18% at Melbourne), despite that the most common type of funding at Purdue is associated with teaching of some kind. Responses to these questions are presented in Tables A20 and A21.

Eighteen respondents at Purdue and 24 at Melbourne provided written responses to the request to indicate other departmental admissions requirements or other aspects of the candidate file that were considered important. Responses were grouped into seven categories, as follows: fine arts requirements such as design portfolios; evidence of proficiency in languages other than English; the status or rank of the applicant's undergraduate institution; interviews or conversations conducted via Skype or telephone; clinical work or experience; and, finally, evidence from other tests, such as the Graduate Management Admissions Test (GMAT) and the Graduate Australian Medical School Admissions Test (GAMSAT).

Funding

Local Departmental and Program Standards

The questions in Section VI asked respondents to consider whether local program standards differed from those of the institution with respect to admissions and funding. With respect to admissions, more than one third of the respondents at Purdue indicated that local program English-language proficiency standards were either *much higher* or *higher* than those set by the university. With respect to funding, 50% of the respondents indicated that English-language proficiency standards were *much higher* or *higher* than those set by the university. When considering applicants for teaching assistantships, 39% of the respondents indicated that local program standards were *much higher* or *higher* for applicants being considered for fellowships or scholarships, 33% indicated that standards were *much higher* or *higher* for applicants being considered for research assistantships, and 21% indicated standards *much higher* or *higher* for applicants admitted without funding (Table A23).

Smaller percentages of respondents indicated *much higher* or *higher* local program standards at Melbourne. Nevertheless, 24% indicated that local standards were *much higher* or *higher* for admissions, and 23% reported *much higher* or *higher* standards for fellowships or scholarships, 21% for teaching assistantships, 20% for students admitted without funding, and 19% for research assistantships (Table A24).

Only 6% of the respondents at Purdue and Melbourne, respectively, indicated that they had ever made an appeal for admission of a candidate with a lower-than-required English-language proficiency score (Table A25).

English-Language Testing Methods

Section VII asked respondents to indicate their impressions (favorable vs. not favorable) of different language-testing methods. At both institutions, testing methods evoked similar sets of favorable and unfavorable responses.

At Purdue and at Melbourne, 98% of the respondents were favorable with respect to oral interviews for speaking, and at least 96% were favorable with respect to human scoring for speaking. Human scoring for essays was considered favorable by similarly high proportions of respondents at both institutions. Performance tests for speaking were likewise considered favorably by the vast majority of respondents at each university. Considered together, respondents were more favorable toward any testing method that involved a human being.

The next set of methods that grouped together were those associated with multiple-choice questions and semidirect methods for testing speaking. Percentages of respondents indicating a favorable response to these methods ranged from 59% to 66% across institutions.

The smallest percentages of respondents having a favorable response were associated with computer scoring for essays and speaking. More than 70% of respondents at both institutions indicated unfavorable responses to computer scoring for essays and for computer scoring for speaking (Tables A26 and A27).

Views/Preferences for Current Tests

A discussion follows.

Familiarity

Section VIII asked respondents to indicate familiarity with the TOEFL test, IELTS, and PTE. Because pilot respondents at ECU indicated that they were more familiar with the PBT and/or CBT, as opposed to the iBT, all three TOEFL versions were included in the final form of the survey. In addition, other pilot respondents commented that they were familiar with a particular scale rather than a particular test, so we included the scales in parentheses.

When asked to indicate familiarity with the tests accepted for admissions, what is most striking at both institutions are the percentages of respondents reporting that they were not familiar with any of the tests. This is all the more noteworthy given the voluntary nature of the survey and the probability (mentioned earlier) that those who completed it had greater concern and/or awareness about language issues than those who did not. At Purdue, the percentage reporting that they were not familiar with any version of the language proficiency tests listed ranged from 31% for the TOEFL PBT to 69% for IELTS and 75% for the PTE. While the higher percentages of respondents reporting that they were not familiar with the IELTS and PTE are not surprising, given that the IELTS has only been accepted for 4 years and the PTE for only 2, it is interesting that the percentages reporting lack of familiarity with the newer versions of the TOEFL (CBT and iBT) approached 50%, indicating a significant lag in diffusion of information, despite massive ETS efforts at dissemination in recent years.

Around 25% of respondents at Purdue indicated that they were *somewhat familiar* with the three versions of the TOEFL test. The largest percentage (27%) of respondents reporting that they were *familiar* or *very familiar* with a version of the TOEFL test was associated with the TOEFL PBT; only 14% reported being *familiar* or *very familiar* with the CBT, whereas 16% reported being *familiar* or *very familiar* with the iBT. The percentage of respondents indicating that they were *familiar* or *very familiar* with the IELTS was far lower, at 4%. Less than 1% of respondents indicated familiarity with the PTE (Table A28).

At Melbourne, again, the percentages reporting that they were not familiar with language-proficiency tests were large. Those reporting *not familiar* with a particular test was smallest for the IELTS (25%), as might be expected, given its currency in Australia, followed by 52% for the TOEFL PBT, 48% for the TOEFL CBT, 35% for the TOEFL iBT, and 61% for the PTE. Nineteen percent of the respondents at Melbourne indicated that they were only *somewhat familiar* with the IELTS, and, for the three versions of the TOEFL test, the percentages of respondents who chose *somewhat familiar* ranged from 12% for the iBT to 19% for the PBT. The patterns of response with respect to being either *familiar* or *very familiar* with language proficiency tests were similar to the percentages at Purdue, with 25% indicating one of these categories for the IELTS but much smaller percentages for the CBT and iBT versions of the TOEFL test (<1–5%) and the greatest familiarity being associated with the PBT (12%). These responses are reported in Table A29.

Thus, at both institutions, those unfamiliar with any of the English-language tests used for admission purposes far outnumbered those reporting familiarity. The PTE, unsurprisingly, given its recent appearance, was the least familiar to respondents at both universities, confirming that raising awareness of particular tests among users takes considerable time.

When asked to indicate preference for a particular English-language proficiency test, respondents selecting *no preference* were by far the largest percentage at both institutions (58% at Purdue and 46% at Melbourne). Respondents' lack of preference was completely unexpected, as we believed that there would be a strong preference for the TOEFL iBT at Purdue and the IELTS at Melbourne, reflecting the more established status of TOEFL tests in the United States and IELTS in Australia; however, only 10% of the respondents at Purdue selected the TOEFL iBT, and only 15% of the respondents at Melbourne selected the IELTS.

Also noteworthy is that, at Purdue, a slightly larger percentage of the respondents indicated a preference for the TOEFL PBT (15%) as compared to the iBT (10%; Table A30 for both institutions).

Reasons for Preference

The next series of questions asked respondents to indicate reasons for their preference. As we were expecting strong preferences at both institutions, we listed a series of reasons we thought might be likely and asked respondents to indicate the extent to which they agreed or disagreed with these reasons. It is important to remember that the percentage of respondents actually indicating a preference was small at both institutions (only 27% at Purdue and 21% at Melbourne). Unfortunately, the way we set up this question makes it impossible to determine which test they were referring to when they answered these questions, and our assumption was that respondents at Purdue would be referring to the TOEFL iBT and respondents at Melbourne would be referring to the IELTS. However, as stated earlier, expected preferences were small at both institutions. These qualifications make it difficult to attribute much reliable information to the following series of questions.

The largest percentages of respondents at Purdue selected *strongly agree* or *agree* for (a) familiarity with the test (65%), (b) tradition of use at my institution (61%), (c) general reputation of the test (61%), and (d) the ability of the test to represent English-language proficiency (58%). Of those who indicated a preference at Melbourne, the largest percentages selected *strongly agree* or *agree* for (a) general reputation of the test (81%), (b) the ability of the test to represent English-language proficiency (68%), (c) familiarity with the test (61%), and (d) measurement precision and reliability (56%; Tables A31 and A32).

Reasons for No Preference

More interesting and revealing are the respondents' written responses to the request for an explanation of why they had no preference. One hundred five of the respondents at Purdue and 98 of the respondents at Melbourne provided a written response to this request. These responses were classified into eight categories. By far the most common category of response at both institutions was that of no preference for any test owing to lack of familiarity, with 59% (62) at Purdue and 63% (62) at Melbourne classified as within this category. Most of these written responses stated simply, *lack of familiarity, I don't know enough about them,* or *I have no knowledge of these tests.*

Eleven percent of the responses at both institutions (11 at Purdue and 12 at Melbourne) indicated no preference because of familiarity with only one of the available English-language tests. As one respondent at Purdue stated, "As I have only dealt with the iBT, I can't say whether the others are preferable or not." Another at Melbourne provided a parallel: "I cannot compare, I have only done the IELTS."

Lack of credibility was given as a reason for lack of preference for any test by 6% of respondents at Purdue and 9% at Melbourne. One respondent at Purdue stated, "I don't trust any of their results. I have had students with high scores; nevertheless, I have had to write their theses for them." Another stated, "As far as I can tell, they are all worthless." At Melbourne, a respondent summed up such sentiments nicely by stating, "They are all rorted."[5] A more measured response was provided in the following: "The IELTS and the TOEFL have their limitations and a better system needs to be developed for evaluating a student's English proficiency."

Five or fewer responses across institutions gave the following reasons for their lack of preference: *applicants meet the minimum, trust the experts, prefer other methods or aspects of admissions file, tests are functionally the same*, and *numbers are meaningless*. With respect to the latter category, the following comment indicates the confusion experienced by some test users in the face of different reporting scales:

> The scales are so different that the score becomes meaningless. I cannot use it. I'm not going to bother to determine what a particular score on a particular test means. With GRE verbal scores there is at least a percentile ranking given.

> If all scores were normalized on a 0–100 scale, then they might be helpful. As it is, I ignore the English proficiency scores because they don't mean anything to me.

Lack of understanding can lead some users to look elsewhere for relevant information:

> The committee has only looked to see if the applicant has the minimum TOEFL scores needed by the university and has instead relied more on previous institutions, GRE/GMAT score, and letters of recommendation. Of course, we are finding that this is inadequate as many of our students struggle once in the program and in the job market due to written and oral communication skills. We just don't have enough information about proficiency tests at this point to make a statement about them but really need to educate ourselves better on this issue.

General Assessment Literacy

A discussion follows.

Knowledgeable About Testing and Assessment

Sixty-six percent of the respondents at Purdue and 46% at Melbourne either strongly disagreed or disagreed with the statement *I am knowledgeable about language testing and assessment*. These results align with respondents' statements concerning their lack of familiarity with language proficiency tests. However, it should be noted that this is one of the items with the largest percentages of no-response values (16% at Purdue and 32% at Melbourne; Table A33). It seems that, from this point on, that many respondents simply opted out of the survey, perhaps because of response fatigue and/or the feeling that they had little further information to contribute.

Interested in Testing and Assessment

Fifty-one percent of the respondents at Purdue and 47% at Melbourne either strongly agreed or agreed with the statement *I am interested in language testing and assessment*, whereas 33% at Purdue and 21% at Melbourne either strongly disagreed or disagreed. Again, this was one of the items with the highest percentages of no-response values at both institutions (Table A34).

If You Were Offered Information About Any of the Three English Admission Tests Mentioned Above, Which of the Following Would Be of Interest?

The majority of the respondents (at least 60%) at both institutions indicated that they would be interested in additional information about test content scoring methods, scores presented as percentiles, benchmark samples of writing and speaking, and samples of individual writing and speaking from the tests (Tables A35 and A36).

Twelve respondents at Purdue and 10 at Melbourne provided a written response to Other. These 22 responses were classified into 10 categories. A sample response is provided for each category, as follows:

- Security (five responses, e.g., "assurance the test taker is the applicant")
- Predictive validity (four responses, e.g., "correlation of the test results with the academic performance or likelihood of degree completion")
- Test comparability (two responses, e.g., "Direct comparison of all three test methods")
- Test–retest reliability (two responses, e.g., "Statistical variation when the same applicant is retested and benchmarks where the same applicant tries tests of several types")
- Descriptive statistics (two responses, e.g., "Score distributions by language group and distributions of scores at peer institutions")
- Aid with interpretation (three responses, e.g., "Scores with descriptions of what might be expected of a person who scores at a particular level along with common expected difficulties for a person with the same score")
- Interaction (two responses, e.g., "Personal discussions in English with the potential applicant")
- Test-taking experience (two responses, e.g., "I would like to take the test so I can have the experience")

- English for specific purposes (one response, e.g., "Evaluation of the use of precise mathematical language. This is quite different from English prose")
- Cutoffs (one response, e.g., "How the pass point was chosen")

Interview Data

In this section, we report interview findings under the thematic categories used for coding the data. The findings are exemplified with extracts from the interviews, which are labeled with the respondent's ID number, disciplinary affiliation, and the number of the turn from which the extract is taken. The extracts are selected to illustrate the range of views expressed by respondents but are confined, for reasons of space, to those that directly reflect our research questions. An attempt has been made to link these views to the survey results where possible.

Admissions Process and the Use of Test Scores

Interviewees were asked about the graduate admissions procedures in their respective departments and exactly where English proficiency figured in selection decisions. Their comments on these issues mirror what was presented for the sample as a whole in Tables A9 – A12, A16, and A17.

Responses identified some commonality across the two institutions in the way international applications were handled and some variation within and across faculties depending on a range of factors, including the numbers of international graduate students applying to the program and whether they were applying for a coursework or research program, as mentioned earlier in this report.

Awareness of and Views About Cutscores or Standards for Entry

On the issue of cutscores for admission, interviewees voiced various opinions. Some expressed uncertainty about what the actual minimum cutscores for the university were, as indicated in the following comment:

> I don't know. I don't know enough about them. Quite frankly, I see those scores and if I jog my memory I can sometimes remember what the minimums are supposed to be. And yet I don't even see the files until our graduate secretary has seen them and a lot of time the graduate committee is forwarding those to me, so they've already been checked for some sort of minimum score so that's why I'm not sure where we stand. (P02, Forestry & Natural Resources: 1258)

Nevertheless, most respondents had an opinion about the level of English at which their students were admitted. The view that current minimum standards on English admissions tests are too low was expressed by a number of interviewees (as was the case with around one third of the survey responses; Tables A9 and A10). The following example is a case in point:

> Yes I'd say so, that perhaps the threshold is too low, at least from the perspective of written English. I don't know how the different proficiency scoring methods score English, written English proficiency, but I would say that the standard is too low for a postgraduate university program. (M10, Medicine: 863)

However, it is unclear whether all of those believing the cutscore to be too low actually understood its meaning. The following comment, for example, suggests that the respondents may be expecting too much of what is intended to be only a minimum English entrance requirement:

> People who score within the accepted range on the IELTS or whatever still come in with really, really marginal language performance and that's been happening so often as to make me wonder what the utility of the tests are. (M11, Optometry: 245)

One respondent who seemed well informed about English entry policies described the situation as follows:

> I think there is a sort of a lack of understanding of what the test does and also I mean in general the way I see the sort of entry level proficiency testing is it's a kind of a base-level and you know if they've got to that level then they're

> going to need to work very hard to manage the course but I think there's not really always that understanding amongst faculties, there's an assumption, OK they've passed the IELTS for a start this sort of notion of pass and fail and therefore they're going to be OK. So I think certainly for staff making the decisions about sort of, well setting the entry levels but also sort of more generally I think it would be useful. (M03, Academic Skills: 263)

The assumption that meeting the English entrance requirement was an indicator of adequate English was also held by more than one third of survey respondents at both universities (Table A9). The same view was reiterated by another interviewee with prior experience under a different academic system:

> Here the minimum requirement is automatically considered as satisfactory and a sufficient requirement so not only have you reached the minimum level, you are deemed to have a satisfactory level. (M04, Applied Linguistics: 357)

There also appear to be misapprehensions about exactly what a language test score is designed to predict, as in the following comment from a Purdue respondent, who seems to be assuming that meeting the university's minimum English requirement should be an accurate predictor not just of language ability but also of academic success:

> We have no shortage of examples of students who meet the university's minimum requirement for TOEFL or whatever exam they took, but when they arrived, they were not able to be successful in the first two semesters the way we'd want them to do. (P05, Earth, Atmospheric, & Planetary Sciences: 1438)

A number of respondents pointed to the tension in the determination of minimum cutscores between the need to maintain adequate levels of English and the dependency on revenue from international students in the program, as follows:

> I think we feel quite a tension between the expectation that we need to get in a certain number of students and that raising the IELTS scores might knock out a pool and also wanting to make sure that students are really equipped to succeed but we also get an awful lot of pressure from students begging to be let in when their English is lower. (M07, Environments: 632)

The view expressed here echoes the results of the survey where it was found that, although *A desire to ensure adequate language proficiency* was perceived by many to be the main reason for the university's current minimum English requirements, more than one third of respondents believed the desire to increase class size or increase revenue to be possible reasons (see Tables A11 and A12). According to Marginson (2011), this tension between the desire for quality and the drive for economic growth is present in the higher education sector of many English-speaking countries. Melbourne interviewees, in particular, seemed acutely aware of this tension and of the potentially negative economic consequences of a drop in student numbers should the English proficiency bar be set too high.

A number of respondents mentioned the need for additional language support for admitted students if entry thresholds remained low:

> At some point you have to set a bar, you know and that point sort of depends I guess on two considerations: how many applications you want to get and how much postadmission support you're willing to provide. (M04, Applied Linguistics: 381)

Others acknowledged that, although current minimum cutscores might be sufficient to prevent failure, limited language proficiency may result in underperformance and high levels of stress:

> I think the benchmark is high enough that it actually wouldn't allow them to fail, it just wouldn't necessarily allow them to do as well as they otherwise could. (M09, Medicine: 772)

Respondents at Purdue from programs with large numbers of applicants appeared to be more confident about setting the minimum English level higher than the university minimum requirement. This is consistent with survey findings where higher percentages of Purdue respondents than those at Melbourne reported raising the language proficiency bar beyond minimum levels (Tables A23 and A24). As one interviewee put it:

> The students who are applying to Purdue from international universities are really top notch. That's what gives us the ability to triage people at a high level because we have such a good international reputation that we're getting very,

> very strong students. So, strong students with mediocre English skills are not making the cut. You've got to be a strong student and have strong English skills because that's who's applying to our program. (P11, Medicinal Chemistry & Molecular Pharmacology: 298)

One respondent at Purdue pointed out that improvement in English proficiency over the course of academic study could not be taken for granted.

> I still think we need a very high level of English proficiency. Part of my attitude used to be that OK, for the first year they're going to struggle and then they're going to catch up. Well, geez, maybe that happens for some students but it's a small percentage, and then the level of catching up, there's different levels. (P07, Medicinal Chemistry & Molecular Pharmacology: 1511)

Another interviewee justified the higher cutscores set by her department on the grounds of value for money, given that all graduate students selected by her department are given funded teaching assistantships for a 5-year period:

> But when we admit, when we are funding them as TAs, we expect a minimum spoken score of I think 29? Or 27? [Interviewer confirms 27] And here is something that I found very difficult as graduate chair to explain to applicants, and the reason is this. Unlike many other departments, we fund every student we admit. And we commit almost 5 years funding provided you have minimum GPA and you go through in a timely manner. (P13, Sociology: 1842)

In sum, levels of awareness about cutscores and their meaning varied widely across the two universities, with some able to justify their decision to set higher language proficiency cutscores than required by their university but also significant numbers erroneously believing that the current minimum requirements indicated more than just an acceptable level of English.[6] It seems that, as one of the participants pointed out, many expect that those reaching the university-stipulated minimum cutoffs are capable of meeting the language challenges of graduate study without additional support. Others mistakenly interpret English proficiency scores as predictors of academic performance. These inflated expectations may explain the dismay expressed by some interviewees when finding some of their students struggling to meet academic demands.

Place of English Proficiency in the Process of Selecting Students

Most interviewees indicated that once it had been established that university or local program requirements for English proficiency were met, they did not give further consideration to language test scores and instead looked to other sources of evidence:

> We look at academic transcript, evidence of postgrad study, referee's reports — we get three referee's reports and occasionally some people write, get a report from the referee which has affected them getting in the course in the area of interest they want to study — and communication skills. But essentially they're not going to be offered interview unless their English scores meet the entry requirements full stop. That's a definite hurdle. So once the English requirements are OK then they need all the rest of the other stuff. (M08, Dentistry: 684)

Others reported assessing academic writing proficiency on the basis of various elements in the candidate file:

> With the master's student who I was co-supervising, the main supervisor would have looked at her previous success in other subjects and I would imagine would have looked at some of her written work to know what level of writing she was capable of doing. (M06, Land & Environment: 513)

As for how the different elements of the file were weighted (see Tables A20 and A21), one interviewee reported giving more credence to the TOEFL test than to a student's prior academic record, expressing skepticism about other sources of information:

> Yes, very much. In fact there are times when I pay more attention to the TOEFL subscores than their GPA. The TOEFL is a more level comparison across these things than GPA. . . . I don't believe what they do as an undergrad is the best predictor of their ability as a researcher. (P03, Industrial Engineering: 1293)

By contrast, a number of interviewees gave reasons for *not* giving too much weight to language proficiency in the selection process after the minimum requirements were met. One was aware, based on his own experience, of the difficulties faced by nonnative speakers:

> Having worked overseas myself, I mean I worked in France for six years as well as in the U.S., so I'm used to working in multilingual environments so I'm well aware of difficulties that people who aren't native language speakers have because I went through that myself. In France. So I like to think I'm reasonably understanding and accommodating of language issues. (M02, Engineering: 177)

One respondent pointed to different sources of information about language proficiency in the student file but found problems with all of them:

> The statement of purpose is a great place, although it's easy to get that coached. It's still a good way to see and understand their culture and their mindset about what they think grad school is about. And again, I can't hear them, I don't know if that's something that took them two hours and a cup of coffee or three weeks and fourteen revisions. But I also know that I have very good friends that are exceptionally shy in an oral interview, so I'm not willing to assume whether or not can start waxing poetic in front of your professor when your career depends on you answering well and you are still getting used to driving on the wrong side of the road or whatever it is. I think the listening comprehension, I think the speaking and writing are absolutely critical. I just don't know how we can do a finely tuned evaluation before they get here. (P03, Industrial Engineering: 1295)

It is interesting that this respondent seems to have dismissed (or to be unaware of) the possibility of assessing command of the different English subskills via one of the standardized language tests used for admissions purposes. The English proficiency score, in other words, is treated as a given, a tendency also noted by O'Loughlin (2011), rather than as a potential source of detailed information that can be carefully scrutinized and judiciously weighed against other factors. We may speculate that this is true for many respondents, as none of the interviewees mentioned looking at either the overall or subtest scores for English proficiency in making their final determinations, even though this information is routinely included in the candidate file. This underuse of test information resulting from the misunderstanding of and misplaced reliance on cutscores appears in some cases to result in the blaming of the test for unsatisfactory language proficiency, as will be seen later.

One reason for this apparent neglect of test information appears to be the lack of knowledge about the particular English tests used in the university admissions process, as also indicated in the survey data (see Tables A28 and A29). This issue is considered in the following section.

Beliefs About and Attitudes Toward Language Tests Used for Selection

When questioned about the particular language tests used for selection at their respective universities, many respondents confessed to limited knowledge about their content and format, about what the scores mean, and about how the university's minimum cutoffs are set:

> No. I wouldn't know how to assess the results from one of these English proficiency scores. Now I'm familiar with the TOEFL concept, I know that that is a test for testing people's English proficiency but yeah I don't know how to assess the scores, whether certain scores are acceptable or whether they get As, Bs, or Cs I wouldn't know. (M02, Engineering: 216)
>
> It's not clear to me where the numbers for the cutoffs came from. I can guess that someone at a certain level of score, such as the GRE, has a better probability of doing well. It's never perfect. So it's never that if they get a 510 they'll never succeed but if they get a 520 they will always succeed, I know it doesn't work that way and then I have to translate the scores anyway. But 77 sounds like an arbitrary number and 14 being the minimum subscore or something around there I don't even remember. (P03, Industrial Engineering: 1297)

A number of respondents voiced general skepticism about the language tests used for selection purposes at each university, as reported in relation to the survey data. As seen from the following interview extracts, this skepticism was based

on a range of factors, among which are concerns about the limited discriminatory power of language tests at certain score levels and awareness of measurement error and potential threats to validity, such as a coaching or practice effects on test scores:

> I mean the large-scale tests like TOEFL and IELTS are not very fine-grained. I mean they need to measure across a whole range of proficiencies from zero to native speaker basically. (MO4, Applied Linguistics: 363)
> I think the basic problem is that none of those tests are going to give, given the standard errors, assessment of risk and that's, that's usually how I phrase rejections, I mean this student is too risky for us to take. (M05, Finance: 468)

One respondent alluded to the possibility of fake results:

> They might bring a TOEFL or these different other English-language training course scores but I have heard through the grapevine that they can buy those and so I wonder how legitimate is that. (M06, Land & Environment: 549)

The lack of alignment between scores and the observed language ability of their students was commented on by a number of respondents:

> I think from my, from the little I know about the test, I don't think it provides an accurate portrait of someone's ability, I think some students have come in with the bare minimum IELTS but they're actually pretty good in terms of their ability to do all the things they need to do. Others have come in with maybe slightly higher bands in certain things like reading for example and yet their ability to do the things I've mentioned, like oral presentations, group work, understanding is poor. (M07, Environments: 586)
> In terms of how we think about the TOEFL score, we have become less confident in its ability to predict language proficiency, especially among Chinese applicants. They all have perfect GRE Quantitative scores and generate reasonable TOEFL scores — that are above the Grad School cutoff — and yet do not seem very literate. That's my impression. (Purdue presurvey interview)

As for the question of which test was most appropriate for selection purposes, a number felt that this was unimportant (a finding that also emerged from the survey data; see Table A30):

> Do I think it matters? Not if we use appropriate cutscores. I mean we just adjusted the cutscores for the TOEFL because before they weren't really calibrated to the IELTS cutscores. In the end, whether you use as a broad gatekeeping instrument, whether you use TOEFL, IELTS, or the Pearson test, I think they're all much of a muchness really. (MO4, Applied Linguistics: 375)
> I'm most familiar with TOEFL, but I know what the other one is. I don't know that it would make a big difference for us if they used one or the other. (P10, Speech, Language, & Hearing Sciences: 1705)

One respondent had views about the value of ability descriptions or profiles rather than scores but was fairly vague about which test met these needs:

> Um, well certainly I mean it's, when you look at the scores it's broken down into all these different areas that are of interest to us but others, certainly a couple I've seen are just a number, doesn't really indicate what was involved in the test. (M08, Dentistry: 712)

In sum, the interview data confirmed what emerged from the survey results, namely, that participants in the main have little knowledge about the particular English-language tests used for selection and no strong preference for one or another test, despite the long association of each institution with particular brands. Worthy of note is the fact that among the reasons for not caring offered by some respondents, both in the survey and in the interview, was a distrust of the test results. This distrust appeared in some cases to be born of ignorance rather than an informed awareness of the "inevitable uncertainty" associated with test scores (Spolsky, 2005).

Beliefs About and Attitudes Toward Language Tests in General

While some respondents had opinions about the value of certain kinds of testing, others acknowledged their complete ignorance, stating that language testing was outside their field of academic expertise, and they were prepared to accept that others were better qualified in this area:

> I think I can speak for most of the other academic directors I know as well, we have to accept on faith that the test is doing what it is because that's not our, you're sort of deluded that we actually know something about this, we don't. (M05, Finance: 462)
> When I was taking the survey I realized just how woefully ignorant I am of the proficiency testing. I just make a sort of naïve assumption that somebody somewhere knows a hell of a lot more about English-language proficiency than I do and that'll all be taken care of. (M10, Medicine: 859)

Comments like these confirm the findings of previous research about the lack of LAL (Inbar-Lourie, 2008) among those outside the field of language testing. Though this is hardly surprising, it is likely to be a significant barrier to self-education and to appropriate use of these scores in the decision-making process.

Beliefs About and Attitudes Toward Different Modes of Testing

Interviewees' opinions about different test modalities exemplified in the extracts echo the survey results to a large extent (see Tables A26 and A27). There were differing views about multiple-choice tests, with the majority of respondents expressing diffidence but some willing to believe that such a format could be useful and informative:

> I think multiple choice is very, I think it is the weakest testing. . . . And so I just wonder whether you're not selecting for good test takers. (P04 [1], Speech, Language, & Hearing Sciences: 1394)

Those interviewees willing to express an opinion about appropriate formats for language tests were generally in favor of direct assessments of speaking and writing and suspicious about CBT scoring:

> Look I haven't had any experience, I just find, I think there's so much going on with speaking that the computer just can't capture, it's the same with writing, I have reservations about whether they can really capture speaking or verbals and in writing for nuanced meanings, I just can't believe that a computer can capture all that, that would be my biggest reservation, maybe I'm a technophobe. (M13, Applied Linguistics: 1080)
> Your readability score goes up if you shorten your sentence. No I don't like those, I think I would have to say I'm a bit old-fashioned. I think that, again I think you could learn very quickly how to deal with those readabilities. So if you want to avoid getting a green line in word then you never put anything in the past tense. (M12, Zoology: 1014)

These interview responses confirm the distrust of indirect and automated approaches to language testing revealed in the survey responses but are generally indicative of what O'Loughlin (2011) termed "folkloric" beliefs, rather than any real understanding of the methods concerned.

Alternative Testing Methods Proposed

Suggestions about alternatives to current language tests were varied and, in some cases, involved formats that are currently part of the IELTS, TOEFL, and/or PTE:

> I know when I was in grad school, they made us go through an oral exam as soon as we got there. . . . I think that is a good thing. They need to go one-on-one oral interview to check your English and you can't fake that. (P09, Aeronautical & Astrological Engineering: 291)
> Or you know, [get] everyone to write a one-pager about the same topic, like people write about all sorts of things and so that, I mean I think that would be a more useful way to perhaps discriminate. (M01, Nursing: 102)

One respondent advocated a different test for postgraduate students, seemingly unaware that writing is already assessed on all three major English tests and that the threshold could simply be raised:

> I would have a separate test with a higher threshold for research programs given that, as a scientist this is what I tell students all the time, the fundamental role of the scientist is to communicate and we communicate through written expression, that's what we do. Essentially we're like authors, we write, we publish our work in journals and it's there for everyone to see forever so you have got to be able to write well. (M10, Medicine: 889)

In sum, the *alternative* methods proposed by interviewees were often identical to those currently used in one or more of the English admissions tests already in place. Therefore, test users might be considered ripe for further education, and greater understanding of what English proficiency tests consist of would perhaps result in a greater willingness to take note, not just of whether test takers meet minimum requirements, but also of the additional information that available tests can provide about test takers' language proficiency profiles.

Conclusion and Implications

The study surveyed faculty members at two English-speaking universities, one in the United States and the other in Australia, and found broad similarities in their understandings and uses of English test scores for the admission of international graduate students. It appeared that policies and practices surrounding the use of English test scores within these institutions were affected by broader contextual issues that are common across the two countries, such as the higher education sector's dependency on revenue from international students and local considerations relating to teaching and research. The study has drawn attention to these issues as they play out in the contexts of concern.

The study posed three questions. The first was concerned with how English-language proficiency test scores were viewed and used in the graduate admissions process. Results showed that despite widespread concern about levels of English proficiency among international students across the two institutions and the professed importance of English test scores, there was little use of these scores beyond the minimum cutoff and a greater reliance on other sources of evidence when it came to making final selection decisions. The second question asked how familiar test score users were with the English proficiency tests (TOEFL, IELTS, and PTE) currently used for admission of international students. Responses revealed a generally limited knowledge of these tests, although users were, as expected, slightly more familiar with the test most commonly used at their own institution, namely, TOEFL at Purdue and IELTS at Melbourne. Surprisingly, given the status of TOEFL at Purdue and of IELTS at Melbourne, score users at each institution displayed no strong preference for either test as a language selection tool, with most citing lack of familiarity as the main reason for their indifference but some expressing distrust of the information that all such tests provide. The third question probed what institutional score users know and believe about language testing in general and about the testing methods used in different language proficiency instruments. Assessment literacy among the respondents was generally limited. Comments from some revealed a preference for more direct methods of assessment and a distrust of multiple-choice items and automated scoring, but these did not appear to be based on any understanding of current approaches. Nevertheless, many of the study's participants were aware of their ignorance on language testing matters and expressed interest in further information.

This study of understanding and uses of language test scores in the higher education context clearly has a number of limitations, including some flaws with survey design that emerged only after the pilot phase, and the considerable limitations of the sample itself, which comprised volunteers who were not necessarily representative of the targeted population of test users. The fact that many opted out of the survey after beginning it suggests that, as already intimated, the sample who did persevere with answering the questions may represent a more concerned and engaged group of test users than is the norm within each university. That these respondents should profess such limited understandings of language tests, in particular, and language testing, in general, is doubly disturbing, as it suggests even greater levels of ignorance among the wider population of score users. That said, the issues emerging from this study were by and large common across institutions and, in keeping with what has thus far been reported in previous research, suggesting that our findings may not be far off the mark. Moreover, given that the admissions policies and practices at Purdue and Melbourne are not dissimilar to those of other top-tier universities in each country, we may surmise that many of

the results reported are generalizable to other higher education contexts, although such speculation should ideally be supported by further confirmatory research at a range of institutions.

The findings of this study have a number of implications, which are considered briefly in what follows. The first relates to theoretical discussions of test use as being grounded in a social context (e.g., McNamara & Roever, 2006), which are clearly supported in this case. The way language test scores are used in the higher education context, including the levels at which cutscores are set, are clearly at least partly a function of the political and economic climates, including the competition between universities and nations in the education marketplace and the decline of public funding for education. Tests can never be viewed in isolation from such factors, and, while test developers' preoccupation with internal test properties is necessary, the focus on accuracy of scores may be obscured or overridden by other, more pressing social and political concerns that affect the policies and practices surrounding score use in particular institutions.

The formation and implementation of language testing policy are also affected by assessment literacy of the users concerned. Misunderstandings about what English proficiency cutscores represent can, as we have seen, lead to distrust of the associated tests and underuse of the information these tests can provide. A disaffected user is unlikely to engage effectively with policy matters and may not take due account of English-language matters in their decision making.

Though, as noted in our literature review, English-language proficiency is only one of many factors necessary for academic success, it clearly looms large for the respondents in this study, not least because of its known impact on student engagement, completion rates, teaching assignments, and future opportunities. Attending to these issues is best achieved through informed and responsible selection practices, which ensure that users draw appropriate inferences from test scores that are available to them, are aware of the likely consequences of accepting students at various points on the assessment scale, and make decisions in light of available institutional strategies and resources to deal with these consequences.

This leads us back to the debate about who is responsible for test consequences, alluded to in our literature review. It is surely unreasonable to expect that testing agencies be held fully responsible for the limited understandings of and underuse of language test scores that emerged in this study and their potentially negative impact on selection decisions. This is the view taken by Davies (2011), who stated, "They [the language testing profession] accept responsibility for the development of the language tests they work on and for the intended consequences of those tests. But they do not accept responsibility for any unintended consequences. Nor should they" (p. 465).

Although the study signals particular areas where further information could usefully be provided by test developers,[7] if we are to seriously take up with Linn's (1997) idea of distributed responsibility with respect to score use and interpretation, the challenge of ensuring that decision makers in the local context are equipped with adequate and appropriate information must surely be taken up by institutional agents, given that local polices and those responsible for implementing them may be subject to change. There should also be an institutional response when users express questions, concerns, and misinterpretations of language test scores, buttressed by appropriately tailored advice from test developers.

Acknowledgments

We acknowledge and greatly appreciate the support of the TOEFL Research Committee for making this study possible. The completion of this project, however, would not have been possible without the additional support and hard work of administrative, secretarial, and graduate staff at both institutions. For assistance with data collection and analysis, special thanks to Jennifer Redden, Soohwan Park, Nancy Kauper, Donna Baker, Nina Barron-Burke, Kyle McIntosh, and Dennis Koyama at Purdue and to Catriona Fraser, Hyejeong Kim, and Cathleen Benevento at Melbourne.

Notes

1 Unless specified otherwise, in this report, TOEFL means TOEFL iBT and IELTS means IELTS (Academic).

2 In the United States, land grant universities are institutions of higher education originally identified by a state to receive the benefits of the Morrill Acts of 1862 and 1890. These acts granted federal lands and funds in support of institutions that would focus primarily on agriculture, engineering, and science. Most of these institutions have become large public institutions; however, a few, including Cornell and the Massachusetts Institute of Technology, became private schools.

3 In Australia, research degrees are primarily centered on a thesis. Coursework required for research degree programs is relatively limited.
4 Although we had originally planned to include administrative staff respondents as well as graduate faculty, it was discovered from the presurvey interviews that, although many staff members were familiar with language-proficiency tests and scores, their contribution to the admission process was limited to preliminary screening of applications, rather than making decisions about selection.
5 "Dishonest, unreliable, not to be trusted, gamed" (adj.). Australian and New Zealand slang.
6 The IELTS Guide (2009) lists an Overall band score of 6.5 (the minimum cutoff for university entry at both institutions) as "probably acceptable" for linguistically less demanding courses and 7 as "probably acceptable" for linguistically demanding ones (e.g., medicine, law, linguistics, and journalism). Below those levels, it is stated that language study is needed.
7 See O'Loughlin's (2012) recent IELTS-based study for ideas on how such information might be packaged.

References

Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. Oxford, England: Oxford University Press.

Back, K., Davis, D., & Olsen, A. (1996). *Internationalization and higher education: Goals and strategies*. Canberra, Australia: Australian Government Publishing Service.

Baird, B. (2010). *Stronger, simpler, smarter ESOS: Supporting international students—Review of the Education Services for Overseas Students (ESOS) Act, 2000*. Canberra, Australia: Commonwealth of Australia.

Banerjee, J. (2003). *Interpreting and using proficiency test scores* (Unpublished doctoral dissertation). Lancaster University, England.

Birrell, B. (2006). Implications of low English standards among overseas students at Australian universities. *People and Place*, *14*(4), 53–64.

Brice, C. (2005). Coding data in qualitative research on L2 writing: Issues and implications. In P. K. Matsuda, & T. Silva (Eds.), *Second language writing research: Perspectives on the process of knowledge construction* (pp. 159–175). Mahwah, NJ: Lawrence Erlbaum Associates.

Chalhoub-Deville, M., & Turner, C. (2000). What to look for in ESL admission tests: Cambridge certificate exams, IELTS, and TOEFL. *System*, *28*, 523–539.

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

Chow, P. (2012). *U.S. and Australian international student data collection: Key differences and practices.* Retrieved from http://www.iie.org/Research-and-Publications/Project-Atlas

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

Crocker, L. (1997). Editorial: The great validity debate. *Educational Measurement: Issues and Practices*, *16*(2), 4.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, *58*(4), 438–481.

Davies, A. (2011). Ethical codes and unexpected consequences. In G. Fulcher, & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 455–468). New York, NY: Routledge.

Davies, A., & Criper, C. (1988). *ELTS Validation Project report* (Research Report No. 1[i]). London, England: British Council/University of Cambridge Local Examinations Syndicate.

Fisher, M. (1985). Rethinking the "foreign TA problem." In J. D. W. Andrews (Ed.), *Strengthening the teaching assistant faculty*. New directions for teaching and learning series (book 22) (pp. 63–73). San Francisco, CA: Joseey-Bass.

Graham, J. (1987). English proficiency and the prediction of academic success. *TESOL Quarterly*, *23*, 505–521.

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, *11*(3), 255–274.

Haertel, E. H. (1999). Validity arguments for high-stakes testing: In search of the evidence. *Educational Measurement: Issues and Practice*, *18*(4), 5–9.

Harding, L., & Pill, J. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary enquiry. *Language Testing*, *30*(3), 381–402.

IELTS. (2009). *International English language testing system guide*. Cambridge, England: Author.

Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, *25*(3), 385–402.

International Test Commission. (2000). *International guidelines for test use.* Retrieved from http://www.intestcom.org

Kane, M. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practices*, *21*(1), 31–41.

Lane, S., Parke, C., & Stone, C. (1998). A framework for evaluating the consequences of assessment programs. *Educational Measurement: Issues and Practices*, *17*(2), 24–28.

Linn, R. L. (1997). Evaluating the validity of assessments: The consequences of use. *Educational Measurement: Issues and Practices*, *16*(2), 14–16.
Linn, R. L. (1998). Partitioning responsibility for the evaluation of the consequences of assessment programs. *Educational Measurement: Issues and Practices*, *17*(2), 28–30.
McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Malden, MA: Blackwell.
Marginson, S. (2011). It's a long way down: The underlying tensions in the education export industry. *Australian Universities Review*, *53*(2), 21–33.
Mehrens, W. A. (1997). The consequences of consequential validity. *Educational Measurement: Issues and Practices*, *16*(2), 16–18.
Merriam, S. B. (1998). *Case study research in education: A qualitative approach*. San Francisco, CA: Jossey-Bass.
Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York, NY: Macmillan.
Moss, P. (1998). The role of consequences in validity theory. *Educational Measurement: Issues and Practice*, *17*(2), 6–12.
O'Loughlin, K. (2011). The interpretation and use of proficiency test scores in university selection: How valid and ethical are they? *Language Assessment Quarterly*, *8*, 146–160.
O'Loughlin, K. (2012). Developing the assessment literacy of IELTS test users in higher education. *IELTS Research Reports*, *13*(5). Retrieved from http://www.ielts.org/PDF/vol13_Report5.pdf
Popham, W. J. (1997). Consequential validity: Right concern—Wrong concept. *Educational Measurement: Issues and Practices*, *16*(2), 9–13.
Purdue University. (2011). *Purdue data digest.* Retrieved from http://www.purdue.edu/datadigest/2011-12/
Purdue University. (2012). *Oral English proficiency program.* Retrieved from http://www.purdue.edu/oepp
Shepard, L. A. (1997). The centrality of test use and consequences for test validity. *Educational Measurement: Issues and Practices*, *16*(2), 5–24.
Spolsky, B. (2005). *Measured words: The development of objective language testing*. Oxford, England: Oxford University Press.
Stake, R. E. (2000). Case studies. In N. K. Denzin, & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 134–164). London, England: Sage.
Stiggins, R. J. (1991). Assessment literacy. *The Phi Delta Kappan*, *72*(7), 534–539.
Taleporos, E. (1998). Consequential validity: A practitioner's perspective. *Educational Measurement: Issues and Practices*, *17*(2), 20–23.
Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, *29*, 21–36.
University of Melbourne. (2014). *English language requirements.* Retrieved from http://futurestudents.unimelb.edu.au/admissions/entry-requirements/language-requirements
Yin, R. K. (2008). *Case study research: Design and methods* (4th ed.). Thousand Oaks, CA: Sage.

Appendix Tables

Table A1 International Advising and Admissions

	Purdue						Melbourne					
	Yes		No		No response		Yes		No		No response	
	%	(*n*)	%	(*n*)	%	(*n*)	%	(*n*)	%	(*n*)	%	(*n*)
Q 6.4. Have you ever served as an advisor to an international student?	78	(182)	21	(49)	0	(1)	69	(173)	72	(29)	2	(4)
Q 7.5. Do you currently serve as a graduate advisor to any international students?	58	(134)	42	(97)	0	(1)	54	(135)	44	(110)	2	(4)
Q 9.6. Have you ever served on a graduate admissions committee?	73	(169)	27	(62)	0	(1)	47	(117)	51	(128)	2	(4)
Q 10.7. Are you currently serving on a graduate admissions committee?	48	(112)	51	(119)	0	(1)	—	—	—	—	—	—
Q 10.7. Are you currently involved in the graduate admissions process?	—	—	—	—	—	—	29	(73)	69	(172)	2	(4)
Q 11.8. Have you ever discussed the meaning of language- proficiency test scores with colleagues?	68	(157)	32	(74)	0	(1)	51	(128)	47	(117)	2	(4)

Table A2 Graduate Advising Loads: How Many International Students Are You Currently Advising?

	Purdue		Melbourne		Combined	
No.	%	(*n*)	%	(*n*)	%	(*n*)
1	33	(26)	33	(33)	33	(59)
2	24	(19)	21	(21)	22	(40)
3	19	(15)	17	(17)	18	(32)
4	6	(5)	13	(13)	10	(18)
5	3	(2)	5	(5)	4	(7)
6	3	(2)	5	(5)	4	(7)
7	4	(3)	3	(3)	3	(6)
8	3	(2)	1	(1)	2	(3)
10	3	(2)	1	(1)	2	(3)
11	—	—	1	(1)	1	(1)
12	3	(2)	—	—	1	(2)
20	—	—	1	(1)	1	(1)
Mode		*1*		*1*		*1*
Total		*78*		*101*		*179*

Table A3 Degree Types at Purdue and Melbourne

	Purdue		Melbourne	
Program	%	(*n*)	%	(*n*)
Coursework (MA or Pro)	77	(178)	60	(150)
Research (MA or PhD)	91	(210)	82	(204)
No response	*2*	*(5)*	*6*	*(14)*
Total	*100*	*(232)*	*100*	*(249)*

Table A4 Respondents' Familiarity With Degree Type

	Purdue		Melbourne	
Program	%	(*n*)	%	(*n*)
Coursework (MA or Pro)	14	(33)	34	(85)
Research (MA or PhD)	84	(194)	60	(150)
No response	*2*	*(5)*	*6*	*(14)*
Total	*100*	*(232)*	*100*	*(249)*

Table A5 Percentage of International Applicants Likely to Be Admitted

	Purdue		Melbourne	
Range (%)	%	(*n*)	%	(*n*)
<10	20	(46)	14	(36)
10–25	28	(66)	18	(45)
26–50	15	(34)	10	(25)
51–75	6	(15)	4	(10)
>76	5	(11)	5	(13)
Don't know	19	(45)	—	—
No response	*6*	*(15)*	*48*	*(120)*
Total	*100*	*(232)*	*100*	*(249)*

Table A6 Percentage International Students Currently Enrolled

	Purdue		Melbourne	
Range (%)	%	(*n*)	%	(*n*)
<10	9	(20)	12	(30)
10 – 25	19	(43)	18	(44)
26 – 50	31	(72)	16	(39)
51 – 75	14	(33)	12	(20)
>76	13	(29)	7	(17)
Don't know	10	(23)	—	—
No response	*5*	*(12)*	*36*	*(90)*
Total	*100*	*(232)*	*100*	*(249)*

Table A7 Mastery of Content Knowledge in My Program Is Largely Dependent on English-Language Proficiency

	Purdue		Melbourne	
Opinion	%	(*n*)	%	(*n*)
Strongly agree	6	(15)	2	(6)
Disagree	9	(20)	5	(13)
Neither agree nor disagree	8	(18)	9	(23)
Agree	30	(70)	39	(97)
Strongly agree	39	(91)	28	(70)
No response	*8*	*(18)*	*16*	*(40)*
Total	*100*	*(232)*	*100*	*(249)*

Table A8 English Proficiency Is Critical for Students to Perform Adequately in the First Year of My Program

	Purdue		Melbourne	
Opinion	%	(*n*)	%	(*n*)
Strongly agree	6	(15)	1	(3)
Disagree	3	(6)	2	(6)
Neither agree nor disagree	5	(12)	4	(9)
Agree	41	(94)	35	(87)
Strongly agree	38	(87)	42	(104)
No response	*8*	*(18)*	*16*	*(40)*
Total	*100*	*(232)*	*100*	*(249)*

Table A9 I Assume the Language-Proficiency Admissions Requirements Represent

	Purdue		Melbourne	
Opinion	%	(*n*)	%	(*n*)
Minimal English proficiency	52	(121)	30	(74)
Adequate English proficiency	38	(87)	42	(105)
Advanced English proficiency	3	(6)	9	(22)
No response	*8*	*(18)*	*19*	*(48)*
Total	*100*	*(232)*	*100*	*(249)*

Table A10 I Consider the Graduate School's English-Language Proficiency Requirements for Admission

	Purdue		Melbourne	
Opinion	%	(*n*)	%	(*n*)
Too low	32	(75)	34	(84)
Appropriate	38	(88)	35	(86)
Too high	0	(1)	2	(4)
Don't know	22	(50)	11	(27)
No response	*8*	*(18)*	*19*	*(48)*
Total	*100*	*(232)*	*100*	*(249)*

Table A11 I Believe That Purdue University's Rationale for Setting the Current Language-Proficiency Requirements Reflects

	Agree		Disagree		No opinion	
	%	(*n*)	%	(*n*)	%	(*n*)
A desire to ensure adequate language proficiency of all admitted students	65	(139)	24	(51)	11	(24)
A desire to align with peer institutions	60	(128)	7	(15)	33	(71)
A desire to secure a large pool of international applicants	56	(119)	20	(42)	25	(53)
A desire to ensure a high rate of degree completion	52	(112)	26	(56)	21	(46)
A desire to increase revenue	44	(95)	29	(63)	26	(56)
A desire to select only the best graduate students	42	(89)	39	(83)	20	(42)
A desire to increase program and class sizes	36	(78)	32	(69)	31	(67)
Responses	*92*	*(214)*				
No response	*8*	*(18)*				
Total	*100*	*(232)*				

Table A12 I Believe That the University of Melbourne's Rationale for Setting the Current Language-Proficiency Requirements Reflects

	Agree		Disagree		No opinion	
	%	(*n*)	%	(*n*)	%	(*n*)
A desire to ensure adequate language proficiency of all admitted students	71	(142)	22	(44)	7	(15)
A desire to ensure a high rate of degree completion	69	(139)	20	(40)	11	(22)
A desire to secure a large pool of international applicants	58	(117)	25	(50)	17	(34)
A desire to align with peer institutions	55	(111)	11	(23)	33	(67)
A desire to select only the best graduate students	49	(99)	38	(77)	12	(25)
A desire to increase program and class sizes	47	(94)	33	(66)	20	(41)
A desire to increase revenue	34	(68)	19	(39)	20	(40)
Responses	*81*	*(201)*				
No response	*19*	*(48)*				
Total	*100*	*(249)*				

Table A13 Purdue: Does Your Program Offer Funding for International Students in the Form of – –?

	Yes		No		Don't know		No response	
	%	(*n*)	%	(*n*)	%	(*n*)	%	(*n*)
Teaching assistantships	75	(175)	10	(24)	6	(15)	8	(18)
Research assistantships	66	(154)	17	(40)	9	(20)	8	(18)
Fellowships/scholarships	60	(140)	18	(42)	14	(32)	8	(18)

Table A14 Melbourne: Does Your Program Offer Funding for International Students in the Form of – –?

	Yes		No		Don't know		No response	
	%	(*n*)	%	(*n*)	%	(*n*)	%	(*n*)
Fellowships/scholarships	60	(149)	17	(43)	—	—	(23)	(57)
Research assistantships	17	(42)	60	(150)	—	—	(23)	(57)
Teaching assistantships	15	(37)	62	(155)	—	—	(23)	(57)

Table A15 Does Your Program Admit Students Without Funding?

	Purdue		Melbourne	
Response	%	(*n*)	%	(*n*)
Yes	75	(120)	48	(119)
No	10	(71)	29	(73)
Don't know	6	(23)	—	—
No response	*8*	*(18)*	*23*	*(57)*
Total	*99*	*(232)*	*100*	*(249)*

Table A16 Purdue Responses on the Importance of English-Language Proficiency Scores

	Very important/important		Somewhat important/not important		Not required	
	%	(*n*)	%	(*n*)	%	(*n*)
Total/overall English score	86	(174)	13	(26)	1	(2)
English writing score	84	(170)	12	(25)	4	(7)
English reading score	83	(168)	13	(26)	4	(8)
English speaking score	79	(160)	16	(33)	5	(9)
English listening score	79	(159)	15	(30)	6	(13)
Responses	*87*	*(202)*				
No response	*13*	*(30)*				
Total	*100*	*(232)*				

Table A17 Melbourne Responses on the Importance of English-Language Proficiency Scores

	Important		Not important		Not required	
	%	(*n*)	%	(*n*)	%	(*n*)
English reading score	92	(161)	7	(13)	0	(1)
English writing score	91	(159)	8	(15)	0	(1)
Total/overall English score	90	(158)	9	(16)	0	(1)
English listening score	88	(154)	11	(20)	0	(1)
English speaking score	87	(152)	13	(22)	0	(1)
Responses	*70*	*(175)*				
No response	*30*	*(74)*				
Total	*100*	*(249)*				

Table A18 Purdue GRE® Scores Ranked by Percentage Reporting Very Important/Important

	Very important/important		Somewhat important/not important		Not required	
GRE section	%	(*n*)	%	(*n*)	%	(*n*)
Quantitative Reasoning	74	(150)	19	(39)	6	(13)
Total	71	(145)	21	(42)	7	(14)
Verbal Reasoning	71	(144)	22	(45)	6	(13)
Analytical Writing	69	(140)	21	(43)	9	(19)
Responses	*87*	*(202)*				
No response	*13*	*(30)*				
Total	*100*	*(232)*				

Table A19 Melbourne GRE Scores Ranked by Percentage Reporting Very Important/Important

	Very important/important		Somewhat important/not important		Not required	
GRE section	%	(*n*)	%	(*n*)	%	(*n*)
Quantitative Reasoning	67	(117)	10	(18)	23	(40)
Total	67	(117)	10	(17)	24	(41)
Verbal Reasoning	65	(114)	10	(18)	25	(43)
Analytical Writing	62	(109)	13	(22)	25	(44)
Responses	*70*	*(175)*				
No response	*74*	*(74)*				
Total	*100*	*(249)*				

Table A20 Purdue Admissions File Components Ranked by Percentage Reporting Very Important/Important

	Important		Not important		Not required	
Component	%	(*n*)	%	(*n*)	%	(*n*)
Transcripts/academic record	88	(178)	11	(23)	0	(1)
Statement of purpose	84	(170)	15	(30)	1	(2)
Letters of recommendation	79	(159)	21	(42)	0	(1)
Research experience	64	(130)	26	(53)	9	(19)
Writing sample/research/proposal	59	(120)	15	(30)	26	(52)
Work experience related to the field of study	53	(108)	36	(72)	11	(22)
Any form of oral interview	50	(101)	16	(32)	34	(69)
Publications	42	(4)	45	(90)	14	(28)
Awards from previous institutions	37	(74)	61	(124)	2	(4)
Teaching experience	22	(45)	63	(128)	14	(29)
Responses	*87*	*(202)*				
No response	*13*	*(30)*				
Total	*100*	*(232)*				

Table A21 Melbourne Admissions File Components Ranked by Percentage Reporting Very Important/Important

	Important		Not important		Not required	
Component	%	(*n*)	%	(*n*)	%	(*n*)
Transcripts/academic record	96	(168)	4	(7)	0	(0)
Any form of oral interview	77	(136)	13	(22)	10	(17)
Writing sample/research/proposal	73	(127)	22	(38)	6	(10)
Awards from previous institutions	65	(114)	3	(56)	3	(5)
Statement of purpose	65	(114)	3	(56)	3	(5)
Research experience	62	(108)	31	(54)	7	(13)
Letters of recommendation	61	(106)	38	(66)	2	(3)
Publications	49	(86)	42	(74)	9	(15)
Work experience related to the field of study	45	(79)	49	(85)	6	(11)
Teaching experience	18	(32)	69	(120)	13	(23)
Responses	*70*	*(175)*				
No response	*30*	*(74)*				
Total	*100*	*(249)*				

Table A22 Is There Any Other Information That You Require for Admission?

	Purdue		Melbourne	
Response	%	(*n*)	%	(*n*)
Yes	9	(20)	10	(26)
No	78	(182)	60	(149)
Responses	*87*	*(202)*	*70*	*(175)*
No response	*13*	*(30)*	*30*	*(74)*
Total	*100*	*(232)*	*100*	*(249)*

Table A23 Purdue Local Standards as Compared to Institutional Requirements

	Much higher		Higher		No difference		Lower		Much lower		
Category	%	(*n*)	%	(*n*)	%	(*n*)	%	(*n*)	%	(*n*)	Response
Admissions	8	(14)	31	(53)	58	(98)	1	(2)	1	(2)	(169)
Teaching assistantships	17	(25)	33	(50)	48	(73)	1	(2)	1	(1)	(151)
Research assistantships	5	(7)	28	(38)	64	(87)	2	(3)	1	(2)	(137)
Fellowships/scholarships	12	(16)	27	(37)	60	(81)	1	(1)	1	(1)	(136)
Students admitted without funding	3	(3)	18	(21)	75	(85)	3	(3)	2	(2)	(114)

Table A24 Melbourne Local Standards as Compared to Institutional Requirements

	Much higher		Higher		No difference		Lower		Much lower		
Category	%	(*n*)	%	(*n*)	%	(*n*)	%	(*n*)	%	(*n*)	Response
Admissions	5	(7)	19	(28)	72	(103)	4	(6)	0	(0)	(144)
Teaching assistantships	5	(3)	16	(10)	76	(48)	3	(2)	0	(0)	(63)
Research assistantships	3	(2)	16	(11)	74	(50)	7	(5)	0	(0)	(68)
Fellowships/scholarships	6	(7)	17	(19)	73	(80)	3	(3)	0	(0)	(109)
Students admitted without funding	4	(4)	16	(17)	77	(82)	3	(3)	0	(0)	(106)

Table A25 Have You Ever Made an Appeal to the University for a Waiver of Minimal Language-Proficiency Requirements for an Applicant Who Had Scores That Were Lower Than the Required Score?

	Purdue		Melbourne	
Category	%	(*n*)	%	(*n*)
Yes	6	(15)	6	(16)
No	80	(185)	63	(156)
No response	*14*	*(32)*	*31*	*(77)*
Total	*100*	*(232)*	*100*	*(249)*

Table A26 Purdue Respondents' Attitudes Toward Testing Methods and Formats Ranked by Percentage Reporting Very Favorable/Favorable

	Favorable		Not favorable	
	%	(*n*)	%	(*n*)
Oral interviews for speaking	98	(194)	2	(4)
Human scoring for speaking	96	(191)	4	(7)
Human scoring for essays	92	(183)	8	(15)
Performance tests (e.g., a presentation) for speaking	92	(182)	8	(16)
Semidirect for speaking (responses recorded, no interviewer)	66	(131)	34	(67)
Multiple choice for listening comprehension	65	(129)	35	(69)
Multiple choice for reading comprehension	65	(128)	35	(70)
Computer scoring for essays	28	(56)	72	(142)
Computer scoring for speaking	22	(43)	78	(155)
No response	*15*	*(34)*		
Total	*100*	*(232)*		

Table A27 Melbourne Respondents' Attitudes Toward Testing Methods and Formats Ranked by Percentage Reporting Very Favorable/Favorable

	Favorable		Not favorable	
	%	(*n*)	%	(*n*)
Human scoring for speaking	98	(167)	2	(3)
Oral interviews for speaking	98	(166)	2	(4)
Human scoring for essays	98	(166)	2	(4)
Performance tests (e.g., a presentation) for speaking	89	(151)	11	(19)
Multiple choice for listening comprehension	62	(106)	38	(64)
Semidirect for speaking (responses recorded, no interviewer)	61	(104)	29	(66)
Multiple choice for reading comprehension	59	(101)	41	(69)
Computer scoring for essays	26	(44)	74	(126)
Computer scoring for speaking	24	(40)	76	(130)
No response	*32*	*(79)*		
Total	*100*	*(249)*		

Table A28 Purdue: Familiarity With English-Language Proficiency Tests

	Not familiar		Somewhat familiar		Familiar		Very familiar	
	%	(*n*)	%	(*n*)	%	(*n*)	%	(*n*)
1. TOEFL PBT (310–677)	31	(72)	26	(62)	17	(39)	10	(24)
2. TOEFL CBT (0–300)	45	(105)	26	(60)	9	(21)	5	(11)
3. TOEFL iBT (0–120)	44	(103)	25	(57)	11	(26)	5	(11)
4. IELTS (Bands 0–9)	69	(160)	11	(26)	3	(8)	1	(3)
5. PTE (10–90)	75	(174)	9	(20)	<1	(1)	<1	(2)
Responses	*85*	*(197)*						
No response	*15*	*(35)*						
Total	*232*	*(100)*						

Table A29 Melbourne: Familiarity With English-Language Proficiency Tests

	Not familiar		Somewhat familiar		Familiar		Very familiar	
	%	(*n*)	%	(*n*)	%	(*n*)	%	(*n*)
1. TOEFL PBT (310–677)	35	(87)	19	(46)	12	(30)	<1	(7)
2. TOEFL CBT (0–300)	48	(119)	14	(35)	5	(12)	<1	(4)
3. TOEFL iBT (0–120)	52	(129)	12	(31)	<1	(6)	<1	(4)
4. IELTS (Bands 0–9)	25	(62)	19	(46)	14	(34)	11	(11)
5. PTE (10–90)	61	(152)	4	(10)	<1	(8)	0	(0)
Responses	*68*	*(170)*						
No response	*32*	*(79)*						
Total	*100*	*(249)*						

Table A30 Which English-Language Proficiency Test Do You Prefer?

	Purdue		Melbourne	
	%	(*n*)	%	(*n*)
1. TOEFL PBT	15	(35)	5	(13)
2. TOEFL CBT	2	(5)	0	(0)
3. TOEFL iBT	10	(23)	<1	(2)
4. IELTS	0	(0)	16	(38)
5. PTE	0	(0)	<1	(2)
6. No preference	58	(134)	49	(115)
No response	*15*	*(35)*	*34*	*(79)*
Total	*100*	*(232)*	*100*	*(249)*

Table A31 Purdue: Preference for the Test Based On

	Agree		Disagree		No opinion	
	%	(*n*)	%	(*n*)	%	(*n*)
Familiarity with the test	65	(40)	5	(3)	31	(19)
Tradition of use at my institution	61	(38)	5	(3)	34	(21)
General reputation of the test	61	(38)	5	(3)	34	(21)
Ability of the test to represent English-language proficiency	58	(36)	15	(9)	27	(17)
Measurement precision/reliability	47	(29)	6	(4)	47	(29)
Security	45	(28)	10	(6)	45	(28)
Scoring method (human/automated)	45	(28)	5	(3)	50	(31)
Method of testing reading	40	(25)	5	(3)	55	(34)
Content of the test	39	(24)	10	(6)	52	(32)
Mode of delivery (paper and pencil/online)	39	(24)	6	(4)	55	(34)
Method of testing listening	35	(22)	6	(4)	58	(36)
Method of testing speaking	35	(22)	5	(3)	60	(37)
Method of testing writing	34	(21)	5	(3)	61	(38)
Peer institutions	31	(19)	8	(5)	61	(38)
Available research	27	(17)	8	(5)	65	(40)
Ability of the test to predict academic success	27	(17)	24	(15)	48	(30)
Turnaround time for results	21	(13)	23	(14)	56	(35)
Cost	13	(8)	31	(19)	56	(35)
Other	6	(4)	3	(2)	90	(56)
Response	*27*	*(62)*				
No response	*73*	*(170)*				
Total	*100*	*(232)*				

Table A32 Melbourne: Preference for the Test Based On

	Agree		Disagree		No opinion	
	%	(*n*)	%	(*n*)	%	(*n*)
General reputation of the test	81	(43)	0	(0)	19	(10)
Ability of the test to represent English-language proficiency	68	(42)	5	(3)	15	(9)
Familiarity with the test	61	(38)	3	(2)	23	(14)
Measurement precision/reliability	56	(35)	3	(2)	27	(17)
Method of testing writing	52	(32)	2	(1)	34	(21)
Method of testing speaking	52	(32)	2	(1)	34	(21)
Method of testing reading	50	(31)	2	(1)	35	(22)
Method of testing listening	50	(31)	2	(1)	35	(22)
Tradition of use at my institution	50	(31)	11	(7)	26	(16)
Content of the test	48	(30)	6	(4)	32	(20)
Available research	47	(29)	3	(2)	37	(23)
Mode of delivery (paper and pencil/online)	47	(29)	3	(2)	37	(23)
Ability of the test to predict academic success	51	(27)	17	(9)	32	(17)
Scoring method (human/automated)	42	(26)	5	(3)	41	(25)
Peer institutions	39	(24)	6	(4)	42	(26)
Security	27	(17)	10	(6)	50	(31)
Turnaround time for results	27	(17)	10	(6)	50	(31)
Cost	11	(7)	15	(9)	61	(38)
Other	10	(4)	5	(2)	85	(33)
No response	*78*	*(195)*				
Total	*100*	*(249)*				

Table A33 I Am Knowledgeable About Language Testing and Assessment

	Purdue		Melbourne	
Opinion	%	(*n*)	%	(*n*)
Strongly disagree	22	(51)	15	(38)
Disagree	40	(92)	31	(77)
Agree	19	(43)	15	(37)
Strongly agree	4	(10)	7	(17)
No response	*16*	*(36)*	*32*	*(80)*
Total	*100*	*(232)*	*100*	*(249)*

Table A34 I Am Interested in Language Testing and Assessment

	Purdue		Melbourne	
Opinion	%	(*n*)	%	(*n*)
Strongly disagree	9	(22)	3	(7)
Disagree	24	(55)	18	(45)
Agree	44	(102)	41	(101)
Strongly agree	7	(17)	6	(16)
No response	*16*	*(36)*	*32*	*(80)*
Total	*100*	*(232)*	*100*	*(249)*

Table A35 Purdue: If You Were Offered Information About Any of the Three English Admission Tests Mentioned Above, Which of the Following Would Be of Interest?

	Interested		Not interested	
	%	(*n*)	%	(*n*)
Description of test content	80	(157)	20	(39)
Clarification of scoring method	77	(151)	23	(45)
Scores presented as percentiles	77	(150)	24	(46)
Benchmark samples of writing and speaking at each score level	76	(148)	23	(46)
Samples of individual writing/speaking from the test	69	(135)	31	(61)
Other	15	(29)	85	(166)
No response	*16*	*(36)*		
Total	*100*	*(232)*		

Table A36 Melbourne: If You Were Offered Information About Any of the Three English Admission Tests Mentioned Above, Which of the Following Would Be of Interest?

	Interested		Not interested	
	%	(*n*)	%	(*n*)
Description of test content	81	(136)	19	(31)
Clarification of scoring method	79	(132)	21	(36)
Scores presented as percentiles	64	(136)	36	(61)
Benchmark samples of writing and speaking at each score level	81	(118)	19	(32)
Samples of individual writing/speaking from the test	70	(11)	30	(50)
Other	11	(29)	90	(94)
No response	*16*	*(36)*		
Total	*100*	*(232)*		

Suggested citation:

Ginther, A., & Elder, C. (2014). *A comparative investigation into understandings and uses of the TOEFL iBT® Test, the International English Language Testing Service (Academic) Test, and the Pearson Test of English for Graduate Admissions in the United States and Australia: A case study of two university contexts* (TOEFL iBT Research Report No. 24, ETS Research Report No. RR-14-44). Princeton, NJ: Educational Testing Service. doi:10.1002/ets2.12037

Action Editor: Gary Ockey

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