

## Maintaining the connection between test and context: A language test for university admission

John Pill  
American University of Beirut

This paper reflects on a review of an existing English language examination for admission to an English-medium university in a non-English-dominant context. Studying how well an established test sits in its present context may highlight environmental changes causing gaps and points of friction. Such an evaluation therefore provides a baseline understanding from which to move forward.

From the 1960s to 1980s, experts developed an examination for applicants to the American University of Beirut that was similar to the Test of English as a Foreign Language (TOEFL) of that time. The AUB English Entrance Examination has remained relatively unchanged since then. Concern about its effectiveness prompted a recent review, providing an opportunity to study consequences of employing a test not fully adapted to its current use.

The review found differences in what is/was viewed as appropriate test format and content, and in definitions of language proficiency. It also noted unwarranted assumptions made about comparability of results from different tests. Current language practices at the university, in the region and in the globalized workplace where graduates subsequently seek employment are different from those assumed when the test was first developed. This indicates the need for test revision and, for example, the potential benefit of developing an institutional language policy.

**Key words:** English as the medium of instruction (EMI), test evaluation, testing English for academic purposes (EAP), university admission test

## Introduction

This paper considers aspects of a review of an English language test in use for many years at an English-medium university located in a non-English-dominant environment. It provokes reflection on two main topics: the perceived need for English language proficiency tests to be taken by some applicants to English-medium universities, and implications arising from the review for the particular test's validity in its context. To start, background information is provided about university admission tests and language test validation.

In a chapter on the use of large-scale tests of English for academic purposes (EAP) in university admissions, Xi, Bridgeman and Wendler (2014) show how consideration of applicants' language proficiency was taken up in various contexts during the 20th century to become common practice, so that 'Typically, a minimum cut score on English language tests is established to screen applicants who are non-native speakers of English' (pp. 318–319). A check of admission requirements for tertiary institutions in English-dominant contexts indicates that setting a minimum standard of English language proficiency through test scores is usual. This requirement is applied mainly to 'international' applicants, that is, those from other countries, where English may not be the dominant language, and not to 'local' applicants, who are assumed to have satisfactory language skills as a corollary of their completion of the national, or local, secondary education curriculum and examinations presumed to be delivered in English. This distinction has never applied properly to institutions that use English as the medium of instruction and are located in *non*-English-dominant contexts (e.g., the institution described here), but such contexts have not received much attention in research. In present times, *any* supposed distinction between 'international' and 'local' applicants or students is losing clarity and relevance in English-dominant contexts as well (Read, 2015). This blurring is due to the increasing mobility of students in a globalized market for education and to the growing use of English as a lingua franca, both in general and as the chosen medium of instruction in institutions (or for particular programs or courses) which previously used another, local, language. Implications of this blurring are considered in this paper, along with further current challenges for EAP admission test development and research.

The case for the evaluation of language tests through 'specialized practices called "validation"' is made by Chapelle and Voss (2014): 'When ... a candidate for university admissions, for example, obtains a score on a language test that is used to decide his or her future, one should ask what

kind of evaluation process the test itself has been put through to assess the adequacy of the score it produces for making such decisions' (p. 1079). This paper's purpose is not to provide a comprehensive validity argument for the test in question. Nevertheless, the review indicates that a systematic approach to test validation will most likely prove useful as the test is revised. Evaluation is viewed as a means to describe a current situation in order to effect change.

The paper has four main sections. The first provides contextual background for the English Entrance Examination of the American University of Beirut. The second briefly describes the data collected in the evaluation of this test, while the third presents findings and observations in terms of what has changed in the test's context. The fourth section considers implications of these findings and observations both for a revision of the existing test and for the institution more broadly.

### **The test and its context**

The American University of Beirut (AUB) is a private tertiary institution originally established as the Syrian Protestant College by American missionaries in 1866. It currently provides a liberal arts education based on the American model to a student body of over 8000 and has about 800 teaching faculty. The university's medium of instruction is English. Although in the early years Arabic was used, English became the official language of instruction in the Literary Department in 1879, and the Medical Department followed a few years later (Anderson, 2011, p. 40). The university has admitted women since 1922, and admission is not restricted by a student's religion, sect, race or nationality. Lebanese citizenship is held by 77% of the present-day student body, while students '[represent] more than 85 countries' (*Facts and figures 2015, 2014*, p. 5). The university's mission is 'to provide excellence in education, to participate in the advancement of knowledge through research, and to serve the peoples of the Middle East and beyond' (see [www.aub.edu.lb](http://www.aub.edu.lb)).

The current university catalogues (*Graduate Catalogue 2015–16*, n.d.; *Undergraduate Catalogue 2015–16*, n.d.) demand that admitted students meet an English language proficiency requirement that is defined using scores obtained on language proficiency tests – namely, the Test of English as Foreign Language (TOEFL), the AUB English Entrance Examination (AUB-EN) and the International English Language Testing System (IELTS) – and on other tests taken to be equivalent.

## The English Entrance Examination

Understanding the conception and development over time of a language test for university admission has been found useful in other contexts (e.g., Davidson & Cho, 2001). This paper does not seek to present a full history of the AUB English Entrance Examination (henceforth AUB-EN) but draws on information about the test found in publically available theses and research papers.

In 1964, the Center for English Language Research and Teaching was set up at AUB with assistance from the US Agency for International Development ('English Teaching Center in Beirut', 1967; Lee, 1971). One initiative was a graduate program leading to diploma or master's qualifications in teaching English as a foreign language (Kreidler & Pedtke, 1970, p. 9); master's theses subsequently submitted at AUB include studies involving the AUB-EN. For example, one thesis reports how the previous system of accepting students on recommendation from certifying schools 'was discontinued by vote of the University Senate of June 20, 1960', and 'An English entrance examination was (and still is) a pre-requisite for admission except for graduates of American high schools' (Sukhun, 1978, pp. 6–7). In her thesis presenting research comparing AUB-EN results with those from a cloze test and a writing test, Shikhani (1982) states, 'At the American University of Beirut, the Office of Tests and Measurements introduced an English test in 1962 for the purpose of evaluating the language proficiency of applicants' (pp. 6–7). Haddad (1995, p. 2) indicates that the Office of Tests and Measurements (now the Office of Institutional Research and Assessment) was set up specifically to develop this new test.

Publications describing research carried out at AUB in the early 1970s report the use of the AUB-EN as a standardized criterion test against which other variables are measured. Stubbs and Tucker (1974) investigated correlations between results on a cloze test and the AUB-EN, while Scott and Tucker (1974) used the AUB-EN as a pre-/post-treatment measure in error analysis research on students preparing for tertiary-level studies. Scott and Tucker (1974) introduce the test in their study as follows. (According to Haddad, 1995, p. 2, the AUB-EN was referred to as the 'EEE' before 1974; in fact, this name remains much used, including in the current university catalogues.)

The *EEE*, developed and standardized by the Office of Tests and Measurement at AUB, consists of four parts: 1) Structure, 2) Vocabulary, 3) Miscellaneous Tests of Language Aptitude, and 4) Reading Comprehension. The Miscellaneous group of tests consist[s] of the following types of items: multiple-choice items of

syntax, multiple-choice items of contextualized vocabulary, clichés, sentences to proofread, etc. Part IV of the *EEE*, Reading Comprehension, contains two sub-sections: a) sentence equivalents where the examinee is presented with a sentence and five alternatives from which to select the equivalent version, and b) a paragraph of 300-400 words followed by five to ten multiple-choice items to test comprehension. (p. 240)

A paper by Hanania and Shikhani (1986), drawing on Shikhani (1982), describes the 'AUB English test' as 'similar in purpose and content to other known ESL [English as a second language] proficiency tests, such as TOEFL' (p. 98). The test described by the authors appears very similar to the test of a decade earlier outlined by Scott and Tucker (1974) above; it is without direct writing, listening or speaking components. A second link to the TOEFL is made by Hanania and Shikhani (1986, p. 98) in a reference to an internal AUB report from 1983 on the equivalence of the TOEFL and AUB-EN. This report was prepared by George F. Miller, who was director of the Office of Tests and Measurements from 1980 to 1987 and engaged in related work at AUB from 1966. His report relates to the TOEFL format in use from 1976, but it also has a precursor, namely a 'concurrent validity study of the AUB EN and the TOEFL' conducted by Miller and J. E. Lowe in 1966 and therefore based on the original version of the TOEFL. This earlier project is referred to in another master's thesis (Baroudi, 1983, p. 85), which lists around 20 separate research studies relating to the AUB-EN carried out by the Office of Tests and Measurements between the late 1960s and the early 1980s (pp. 82-85). Baroudi's research considers AUB-EN and TOEFL scores as predictors of English grades at AUB (i.e., grades in English language and communication courses). Among several changes suggested for the AUB-EN, he recommends the addition of a listening section, present in the TOEFL from its first instantiation, offered to test takers in 1964 (Spolsky, 1995).

A more recent master's thesis (Haddad, 1995) describes the AUB-EN in use in 1995:

the EN Exam consists of 150 items covering five timed sections: Listening Comprehension, Grammar, Writing, Vocabulary and Reading. Each section consists of 30 items, the time allotted for every section is twenty minutes except for the reading section which requires [forty] minutes. (pp. 2-3)

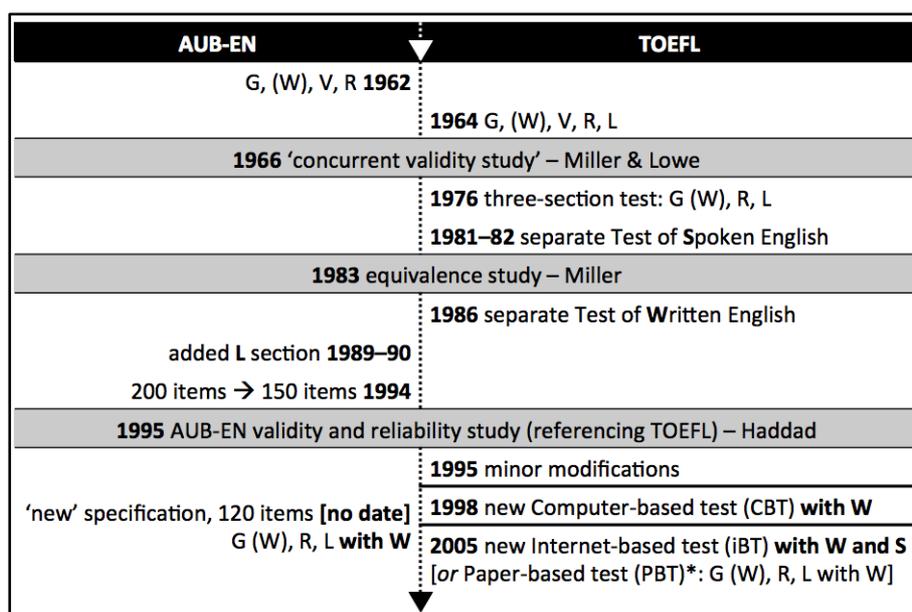
Haddad explains that the listening comprehension section was added in 1989-90 and the number of test items was reduced from 200 to 150 in 1994 (1995, p. 11). The many comparisons with the TOEFL in Haddad's

investigation of the 'validity and reliability of the AUB EN' may indicate the perceived importance of the score concordance and the similarity of format and content between the locally developed test and the American test. It is noteworthy that the two tests were established at more or less the same time – in the early 1960s – but it is not clear whether the first version of the AUB-EN, administered in 1962, could have been designed with reference to the initial TOEFL version, then still under development. The two tests perhaps both represented good practice in the field at that time.

No published documentation has been identified relating to changes made to the AUB-EN as described by Haddad (1995) so that the test matches its present format. The only information available for prospective test takers that describes the current AUB-EN is a three-page document ('New AUB-EN specifications', n.d.). This text is, in fact, a very brief test specification intended for item writers rather than test takers. It describes the current AUB-EN as having three 'objective' sections involving multiple-choice (MCQ) items – listening comprehension (30 items; short conversations, long conversations, talks), written communication (40 items; sentence completion, error recognition, paragraph editing) and reading comprehension (50 items; sentence completion, passage comprehension) – followed by an 'essay part', in which test takers have 30 minutes to write an essay on a given topic. The grading of the essay is described as being holistic, using 'the same criteria established for the TWE' (p. 3), that is, the Test of Written English, which is taken nowadays as part of the TOEFL PBT (paper-based test) only.

Direct reference to current AUB-EN test forms is not made here, as they remain confidential. Little other information relating to the current test was identified during the review. No official sample test material is available and no current test specifications were traceable apart from the document described above, which does not include any sample items. Review participants who had previously contributed to test development as item writers indicated that the most recent test form was prepared four to five years before the review.

Figure 1 summarizes the development of the AUB-EN and the TOEFL, drawing on a brief history of the TOEFL program (Educational Testing Service, 2007, p. 3).



### Legend

G: Grammar, V: Vocabulary, R: Reading, L: Listening, W: Writing, S: Speaking; (W): Writing assessed indirectly.

\* The 1995 TOEFL version plus the Test of Written English, offered only in certain countries.

**Figure 1.** Timeline of major developments in the AUB-EN and TOEFL including three comparative studies undertaken at AUB.

In summary, from the 1960s to the mid-1980s, several research studies were undertaken at AUB relating to innovations in language testing methods. These included using integrative test formats, such as cloze tests, and performance components, such as direct writing tasks, to supplement the discrete-item, multiple-choice format predominant in standardized tests of the period. In parallel, researchers sought to establish and support the validity and reliability of the AUB-EN, in particular through investigation of score concordance with the TOEFL. Haddad's (1995) thesis appears to be the most recent of such studies available; evidence indicates that the test format has been revised at least once since her thesis was submitted.

Likely causes of the apparently reduced focus on developing and monitoring the AUB-EN are not difficult to find. Anderson (2011) summarizes the general situation in the 1980s at AUB: 'During the Lebanese civil war (1975–1989), [AUB's] administrators, faculty, staff, and students struggled to keep the school functioning, with few resources available for building anything new or expanding any of the programs' (p. 184). The war led to a drop in student enrollments; likewise, some faculty members left the country and expertise was not easily replaced. The US government prohibited its citizens from travelling to or working in Lebanon from 1987 to 1997, while many Lebanese

also lived abroad at that time. The return of expatriates during the subsequent period of recovery and normalization was gradual.

### **Reviewing the AUB-EN**

In 2007, a group of experts in the fields of language testing and psychometrics was invited by the provost to visit AUB to advise on the university's use of English language tests for admission and placement. The short consultancy was likely prompted by procedural uncertainty caused by changes in 2005 to the format and scoring of the SAT (a standardized US college admission test required of undergraduate applicants to AUB) and of the TOEFL – the new TOEFL iBT (internet-based test) was introduced from late 2005. This uncertainty may have been augmented by a more general concern that the set of various tests and scores recognized as indicating sufficient English proficiency in students commencing studies at AUB was no longer functioning as coherently as it should.

The findings of the report (Chalhoub-Deville, Brown, & Willse, 2008) are not the focus of this paper, but one of its recommendations was the appointment of 'a professional with appropriate language testing and psychometric skills to manage the English language testing program of the University' (p. 9). In 2012, the provost established a University Committee on English Language with the main role 'to assess and monitor English language standards and policies that affect undergraduate and graduate admission and placement in English Communication Skills courses' (see <http://www.aub.edu.lb/provost/Pages/UCEL.aspx>). A task adopted by the committee was to create the position recommended in the consultancy report. This was approved by university authorities and the author of the current paper took up a post at AUB in September 2014 with 50% of its duties related to 'managing the delivery of quality English language assessment and contributing to the review of English language proficiency standards and requirements' (Department of English, 2013).

An initial task in any new role is to gain an understanding of the existing situation, and it is that process which is presented here as an evaluation. While the author was reviewing the context as newcomer rather than outsider, the two perspectives are similar, perhaps distinguished mainly by the likelihood that the newcomer will have to take action based on the findings whereas the investigator in an external evaluation will not. The induction of a newcomer to the practices of a large and complex institution is inevitably slow. The author recognizes that his understanding remains incomplete in many areas but nevertheless believes that the outsider's point

of view he brings to the situation may have value if it draws attention to aspects that might be taken for granted by those having a longer involvement with the test and its context.

### **Data collection**

During the first (fall) semester of the academic year 2014–15, the author consulted informally with colleagues in the organizational units connected with language testing at AUB and wrote a proposal for a more substantial, focused review to be carried out in the second semester. Its aims were to record existing policy and practice, to establish current and future needs of the institution regarding English language testing, and to set an agenda for the development of processes and policy to support the institution's mission, including an assessment of the resources required to carry this out. The proposal for the review was accepted in January 2015 and the final report (Pill, 2015) was completed at the end of the spring semester.

The review included the collection and evaluation of internal and publically available documentation and research to establish a baseline understanding of the situation. Studies published in the literature on language testing and tertiary admissions were reviewed to find examples of good practice in similar contexts that might inform future initiatives at AUB. The author interviewed stakeholders who work in areas of the university's administration where test scores are used directly for admission decisions (e.g., registrar, director of admissions, director of enrollment management unit) and who deal with the consequences of decisions made based on test scores (e.g., director of the Communication Skills Program regarding the use of test scores for placement into those courses, chairperson of an undergraduate admissions committee). Information was also collected through informal interaction with students, instructors and other members of the university community concerning perceptions of English language testing procedures at AUB. AUB-EN test forms in use were reviewed and routine test administration and assessment procedures were observed. There were practical limitations on the process in terms of resources available (mainly time constraints), and certain issues remain unclear to the author, either due to their complexity or because they seemed self-evident to participants in the review and so were not fully explained.

### **What has changed**

In this section, findings related to the perceived need for a test are introduced first. Later sections present observations on some of the gaps and frictions

between the existing test and various aspects of its context, as observed during the review.

### **The need for a test, and a local test**

The need for applicants to demonstrate a sufficient level of English proficiency for admission to AUB was supported in the review. Participants saw the university's use of English as the medium of instruction for most of its history as a source of prestige and a sign of its international outlook. Some indicated a concern that English language standards had fallen among both students and faculty, perhaps threatening the institution's reputation; the current English language proficiency requirement for admission was perhaps set too low, it was suggested. As comparable institutions in the United States set English entry requirements, it was thought appropriate that AUB did the same. Conforming to the perceived norms of US universities was an important theme. Participants did not raise the issue of whether the 'local'/'international' distinction often made at US institutions (described above) applied in the context of the AUB.

The TOEFL was generally acknowledged as representing the state of the art in English language tests; the IELTS was also mentioned as a test gaining favour among applicants. The AUB-EN was viewed somewhat differently from these tests available worldwide and taken annually over a million times each. It was seen to serve a different test-taker population, as it costs less, is available only on campus in Beirut, and can be offered flexibly (e.g., requiring little lead time) as needs demand. These practical points were also reasons given for the maintenance of the test. Requiring all applicants to take commercial tests offered by external providers might reduce the diversity of the applicant pool, and this would conflict with the university's stated mission. The ideal AUB-EN should reflect the characteristics of the TOEFL iBT and be available on demand and at a price affordable to all potential applicants.

In the main, it was assumed that the AUB-EN and other recognized English language tests were appropriate to their purpose. Overall, knowledge among faculty and administrators about any of these tests – for example, their current formats and content – was limited. A reflection of this is the continued reference in the university catalogues to the TOEFL CBT (computer-based test), which has not been available since 2006. Similarly, the TOEFL PBT (paper-based test) is still presented as the default test version whereas it is now offered in very few countries (although at the time of the review these did include locations likely to be home to some applicants to AUB, such as Syria and Gaza).

To summarize, the AUB-EN is viewed as a necessary element of the current and future application process for AUB. Problems with the current AUB-EN raised in the review are considered below, as a way to establish features expected in a revised test.

### **The language proficiency construct and its representation**

Inspection of test forms indicated that the current AUB-EN represents a rather outdated view of language proficiency. The history of the test outlined above provides an explanation for this. The addition to the test of a listening section at the start of the 1990s and a writing performance task (essay) exemplifies how developments in the field have been recognized. Nevertheless, in comparison with the TOEFL, the current AUB-EN still seems aligned more with the PBT of the 1990s than with today's iBT, and the change in the TOEFL delivery format (from paper- to internet-based) does not fully explain this. The difference relates to the expansion of the construct of language proficiency to include sociolinguistic and strategic competence. While present to some degree in the existing AUB-EN, representing these features is prioritized in the design of contemporary commercial tests.

Starting in the 1980s, the refocusing of the teaching of second languages to encompass *communicative competence* (Canale & Swain, 1980) – that is, knowing how to *use* a language rather than knowing *about* a language – had an impact on language testing, too, although this was gradual and remains incomplete (see Elder, McNamara, Kim, Pill, & Sato, forthcoming; Harding, 2014). The notion informed the development of the original IELTS and its revision in the 1990s ('History of IELTS', n.d.); similarly, the TOEFL iBT, first offered in 2005, reflects a stronger orientation towards communicative competence than the PBT, perhaps partly in reaction to criticism of the PBT in this regard (see, e.g., Spolsky, 1995). Reading and listening sections involve longer, complete texts that provide clear contextualization for interpreting test items. Speaking and writing are compulsory components of the test and they are task based, with test takers responding in simulations of real-world communication. The productive skills are tested directly; test takers complete performance tasks to be assessed by human raters (as well as scored automatically, in the case of written performance). There is no separate, indirect testing of structure and written expression. The AUB-EN, on the other hand, retains a discrete-item approach in some sections. Listening and reading items are based on brief texts (short conversations, sentence completion) for which the context given is minimal, as in the TOEFL PBT. The written communication section (separate from the essay task) provides an indirect test of writing skills through multiple-choice items on sentence completion, error recognition and paragraph editing. Some items studied for

the review seemed to test grammatical accuracy with little regard for the impact of sociolinguistic context.

It is perhaps salutary to note here that the commercial tests being suggested as models for revision of the AUB-EN are themselves dated in many ways. Their own ability to represent adequately the construct of language proficiency as currently conceived has been much criticized. For example, McNamara (2014) views as limited the change in language testing theory and practice over the thirty years since communicative approaches became the paradigm in language teaching. It is perhaps unlikely that large-scale tests will be the source of construct innovation in language testing, a field that is anyway rather cautious when dealing with change.

### **The domain and what is testable**

Another priority for the commercial tests, connected with the inclusion of performance tasks noted above, is to simulate the domain authentically. The specifications for the TOEFL iBT show that its designers seek to represent the linguistic demands that test takers will be required to meet as students undertaking tertiary studies in the medium of English. Test materials simulate texts and tasks encountered by university students. Tasks which integrate skills are included: for example, a test taker may listen to a short simulated lecture on a topic, read a passage presenting a different view on the same topic, and then write a text comparing the views expressed. (This task type is made feasible by the test's computer-mediated delivery.) In contrast, the tasks and content of the AUB-EN do not reflect with the same clarity the domain of language use the test takers are seeking to enter. On review, the AUB-EN appeared to be more an achievement (summative) test reflecting a secondary school curriculum than a proficiency (qualifying) test predicting ability to manage tertiary-level study in English. Obviously, test takers cannot be required to deal with contexts they have not yet encountered – an admission test cannot assume knowledge of specific university procedures or terminology. Nevertheless, it should provide a sufficient simulation of the target language use domain. This may then have positive consequences in terms of washback into preparation practices for the test.

The range of language skills required in a test may have increased not only because of the importance now given to assessing the particular demands made on students at university but also because the scope of those demands themselves has evolved. Teaching and learning practices at today's universities, as well as perceptions of the purpose of higher education, are different from those of 20 years ago. Greater emphasis is placed on, for example, transferrable skills, learning how to learn, self-management,

reflection on learning, student participation in class, group work and peer teaching. An expanded (or re-prioritized) set of language skills is most likely required to perform these behaviours successfully.

A final point, which became clearer to the author when considering the domain and what is testable, is that the AUB context is different from that assumed in commercial, globally available tests. The university is located in a country where multilingualism is the norm, and diversity of language use is notable among its students, faculty and staff. Arabic is the official language of Lebanon, along with French. English plays an important role in education, commerce and tourism and is viewed as an essential skill for career progression (see Shaaban, 1997). AUB students include users of English as a first and as an additional language; many use three languages or more in different contexts on a daily basis. The official statement that English is the language of instruction at AUB does not recognize the several languages and dialects used concurrently and unaffectedly in most spheres of university activity. This state of affairs is itself not a recent change; Arnold's (2014) research shows how issues of linguistic identity and multilingual writing practices were part of university discourse from its earliest days. Even so, neither the AUB-EN nor the commercial tests used for admission to AUB are designed to reflect this situation.

There has been recognition in the field of applied linguistics over the past 10 and more years of the particular characteristics and roles in international English-medium tertiary contexts of different Englishes (e.g., Jenkins, 2011) and of multilingual practices (e.g., Aintzane, Lasagabaster, & Sierra, 2013). This reconceptualization also forces the field of language testing to reconsider the language(s) to be assessed, the construct being measured and the accurate description of the domain (see, e.g., Elder & Davies, 2006; Jenkins & Leung, 2014; Shohamy, 2011). These issues are certainly not new for AUB; however, they may be taken for granted and not discussed explicitly. It is suggested that language use at AUB should be deliberated across the university to inform institutional decisions, and that these decisions might be reflected in the design of the language test used at the institution (see the section 'Where to go from here' below).

### **What can be measured reliably**

A preference, held particularly in US testing contexts (Spolsky, 1995, p. 349), for scoring procedures viewed as eliminating subjectivity led to the almost exclusive use of the MCQ format in the TOEFL PBT. This is still the case for the AUB-EN (except for the essay part). While this format certainly allows for efficient automated marking with bubble sheets and optical character readers,

using MCQ format limits the task types possible in a test, perhaps leading to construct under-representation. Introducing human scoring may compromise such objective measurement, but it admits the use of more task types, particularly performance tasks to test the productive skills, and this may increase a test's authenticity to its domain (see above). Furthermore, it has become feasible for test administrators to carry out routine statistical analysis – using, for example, many-facet Rasch measurement software – which systematically accounts for score differences between human raters when test performances are scored independently by at least two raters (see McNamara, 1996).

### **The comparability of scores on different tests**

In the review, concerns were raised regarding the reliability of AUB-EN scores by some students who had taken the test and by some users of the scores at AUB. These concerns were mainly generated by the availability of scores on other, supposedly comparable, tests for the same applicant/student and the observed lack of equivalence between these scores in the score tables used for admission and placement at AUB. These apparent inconsistencies led test users to doubt the AUB-EN.

The review found that the tables of equivalent scores being used had a complicated history; scores had been linked for pragmatic reasons, with decisions sometimes made based only on small datasets. The first section of this paper notes that 'concurrent validity' studies were carried out for the AUB-EN with the TOEFL early in its development. When the later TOEFL versions (CBT and iBT) were introduced in the early 2000s, tables were published linking scores among the TOEFL versions. Although these tables are no longer to be found on the Educational Testing Service (ETS) website, there was evidence of their use to create expanded score tables at AUB. The addition of the IELTS to the AUB score tables remains incomplete, although, in this case, concordance data that could inform the process are available from a study carried out by the TOEFL provider (ETS, 2010).

The question arises of whether concordances between different language tests and even different versions of the same test are meaningful. This may be the reason the TOEFL score comparison tables are no longer available: the test provider recognizes the substantial differences between the PBT and the iBT in construct and format and therefore also in score meaning. The same reluctance should perhaps be observed about assuming direct links between scores on the AUB-EN and the TOEFL iBT until an empirical study can be undertaken, given the current differences between these tests described above.

This question may be even more significant when equivalence is assumed between scores on tests never designed to measure the same construct. Scores on the writing section of the SAT – an academic aptitude test commonly required for undergraduate admission to US institutions and to AUB – are also given as equivalents on the AUB score table for English language proficiency. The SAT, although a test *in* English, is not a test of language proficiency, and any direct comparison is consequently flawed. Although the point has been understood by those involved in admissions at the university, the situation is difficult to remedy because the erroneous alignment, put in place to streamline admission, has become set in institutional process. Nevertheless, this is an important issue to resolve because the review found that relatively few undergraduate applicants used an English language test (AUB-EN, TOEFL, IELTS) to meet the language proficiency requirement, most doing so with their SAT writing score instead. (A similar situation exists for graduate applicants, as scores on certain sections of common graduate-level aptitude tests are also accepted as indicating an applicant's English language proficiency.)

### **The uses of test scores**

As well as being used for admission to AUB, test scores also serve to place students into the university's Communication Skills Program, which provides students with writing and composition instruction. The program was previously seen as developing students' English language skills but in more recent years its focus has been intentionally changed so that courses are more similar to freshman composition classes common in US universities. Course titles, however, remain unchanged (e.g., Academic English). Students are currently placed into a sequence of courses based on test scores on the SAT, TOEFL or AUB-EN, presented as being equivalent (see above). In addition, students are promoted to a higher-level course on successful completion of its lower-level pre-requisite in the program sequence. In the review, course instructors expressed a belief that the wide variation in student ability apparent in their classes was a symptom of problems with the score equivalence table for placement interacting with the promotion system. Given the inconsistencies described in the previous section, these views are unsurprising. For example, using the AUB-EN, an English language test, to place students into different levels of a sequence of composition courses illustrates a mismatch between the test's purpose (English proficiency) and the program's instructional goals (academic writing skills). Variation in test performance as indicated by an overall test score is not relevant information if the consequent placement is into a program not intended to address directly the construct measured on the test.

It is in the Communication Skills Program and in remedial English courses for students not meeting the English language proficiency requirement that tensions created by inappropriate score use are most noticeable. Nevertheless, here and perhaps elsewhere in the university (e.g., the admissions office), test-score users appear to work to accommodate distortions in the score equivalence tables and mismatches in test purpose and score use. The strains created by a long-established testing process not fitting well with its current environment are therefore somewhat relieved through people familiar with test and context (instructors, program directors, admissions staff) acting as intermediaries. Perhaps unconsciously, they interpret test scores and modify how they teach and advise test takers so that the scores remain serviceable in this 'closed system'.

The purpose of the AUB-EN, and therefore perhaps also the appropriate uses for its scores, may be self-evident to those taking it. However, test scores are used for inappropriate purposes (e.g., placement into composition courses). Providing clearer information to score users about the meaning (and suitable uses) of AUB-EN results is likely to reduce this problem. It is noteworthy how the descriptions of performance (ETS, 2014) given to test takers in their TOEFL score report clearly link their scores to the real-world academic abilities they have demonstrated, making it readily apparent what the test does and does not measure.

In summary, the review indicated that the AUB-EN may no longer sit comfortably within its current context of use. The following section considers the implications of these findings and observations for future development of the test and its use at AUB.

### **Where to go from here**

Given the perceived need to test applicants' English language proficiency at AUB, a basic consideration is to determine the appropriate standard for admission. A recalibration of the entry requirement can be undertaken through a standard-setting exercise, perhaps using materials available from the commercial test providers for this purpose and involving stakeholders from across the university. In addition, it seems necessary to revise the existing AUB-EN to meet current expectations of language tests. To make sense, however, these projects first require reflection on broader issues: the characteristics of AUB as a place where multiple languages are used, and the type of language test that might serve this context.

## Understanding the context

While it seems reasonable to use English proficiency tests to ensure incoming students have sufficient language skills to cope with their studies, it is problematic in this context to agree even on basic issues: who should take a test and who will be exempted, what the minimum standard required should be, and exactly what skills a test should seek to measure.

The AUB community of students, faculty and staff is linguistically diverse; any statement that the institution's medium of instruction is English obviously oversimplifies the interplay of languages evident throughout the university. Distinctions between 'local' and 'international' applicants that may (still) be viewed as meaningful at US institutions certainly do not apply cleanly at AUB. The 'local'/'international' distinction reflects debates about defining 'native' and 'non-native' users of language (see Davies, 2013, and Hulstijn, 2015, for different perspectives). Categories traditionally applied in the field of English language teaching – for example, 'English as a second language' – are found to be unusable in many contexts today (Dewey, 2012). Given the varied backgrounds of applicants to the university (see below), predicting which of them has the necessary language skills will not be straightforward without an appropriate measurement tool.

The challenge is to establish a minimum required level of English language proficiency to enter the institution, while recognizing that students are likely continue to develop their skills through their engagement in the activities of the university. It is suggested that this ongoing linguistic development is more likely in a multilingual context, because of the blurring of divisions that, in monolingual institutions, may push 'international' students out of the mainstream (into 'linguistic ghettos' where their first language is used predominantly). However, it remains difficult to define what the necessary language skills are for entry into a multilingual context such as AUB. While language tests have been relatively effective in assessing grammatical accuracy in one language, they prove less able to capture, for example, a test taker's ability to accommodate easily to the limited repertoire of another speaker in order to reach a successful outcome through the interaction (Jenkins, 2011), or to use more than one language to negotiate a task (Shohamy, 2011). Yet, these skills are arguably very useful in multilingual contexts; they are also skills that seem to characterize successful members of the global community. Judgements of others' communication skills made by people without specialist knowledge of language teaching or linguistics were found to be based on features of performance different from those measured on traditional language tests (Sato, 2014). Perhaps these features – including

demeanour, non-verbal behaviour and content – should be taken more into account when assessing readiness to join the community of learners at a tertiary institution.

### **Information about the test population**

To develop an appropriate test, an understanding of the test population is needed. The review found that relatively few data were available that would constitute a language profile for applicants and students. Indirect data are helpful to varying degrees: citizenship is not useful in indicating language skills (perhaps illustrating better the global commercial connections from Lebanon); details of the secondary school attended and school-leaving qualifications obtained are more useful, indicating language(s) of education (e.g., through the language(s) used in the Lebanese, French or international baccalaureates). Data directly concerning applicants' first language(s) were not available to the author and are presumed not to be collected for admission.

More effective data collection would help deal with important practical issues, for instance, estimating the number of students who need to take a language test for admission if the exemption requirements are stated in a particular way (e.g., based on secondary school qualifications submitted). It would furthermore allow changes in the language profile of applicants and students to be observed over time; this might indicate the need for amendments to a future test, its scope and tasks, or to the appropriate proficiency standard for admission. Collecting more detail regarding language test scores would also help serve students better. Currently, only the overall score is recorded, while the potentially useful profile of scores reported for speaking, listening, reading and writing on the TOEFL and IELTS, and the similar profile on the current AUB-EN, is missing. This profile could inform the curriculum for courses providing remedial English language support. Analysis of language test score profiles may also indicate where particular problems lie in the current score concordance tables for these tests.

More generally, it is difficult to understand fully and to harness the potential of the linguistic capital at AUB without more definite information about it. This diversity is an aspect of the university that often goes unremarked by students and staff perhaps because it is not exceptional in the Lebanese context, but it should be explored further so that the university can develop opportunities it affords.

### **An institutional language policy**

The review demonstrated the need for greater language assessment literacy (Taylor, 2009) among faculty and administrators; providing education in this

area can be a responsibility of the holder of the language testing specialist post at AUB. Knowledge of the format and scope of the commercially available language tests is important if their scores are to be used appropriately; similar knowledge will be required of a revised AUB-EN. Participants in any standard-setting exercise to establish the appropriate language proficiency standard for admission to AUB will need an understanding of the tests recognized for this purpose.

Ideally, decisions about language test use and proficiency standards should be made within a more general framework of policy on the use of language(s) at the university. This may include explicit statements about issues already discussed, justifying the use of English as the medium of instruction, recognizing the linguistic diversity at the institution and the benefits of this for students and graduates, and setting out how AUB seeks to position itself in language terms as an institution in Lebanon, in the region it serves and internationally. The process of developing these statements to the satisfaction of AUB's internal stakeholders will help clarify institutional values in this area. The policy statements can then support consistent and defensible procedures to admit students to the university and place them in courses. Policy can be reviewed regularly to ensure it remains relevant to its context and the needs of the institution.

### **A revised AUB-EN**

As noted above, the TOEFL was developed as a test of English as a *foreign* language for test takers applying to study at US institutions in English; therefore, to some extent, it does not suit the context presented in this paper. Furthermore, standardization for global applicability in the TOEFL and IELTS is achieved by minimizing reference to particular contexts: test constructs are simplified to exclude local practice, and tests thus represent everywhere and nowhere. The tests' content, tasks and criteria are recognizable and appear sufficiently representative of AUB to provide useable information for admission. But, could a local test, a revised AUB-EN, do more? There seems to be scope for a program revising the AUB-EN to create such face validity by operationalizing features that tests with a broader reach cannot. These may include, for example, recognizing the variety of accents in spoken English heard at AUB ('native' and 'non-native'), prioritizing aspects of performance in writing and speaking that are valued in communication by current students, faculty and staff (e.g., communicative effectiveness over grammatical precision), and selecting test content and tasks that reflect the academic, social and administrative tasks new students are likely to encounter. Practical constraints are still necessary, however. For example, in fairness to applicants unfamiliar with the university, knowledge of specific

procedures must not be included, as it cannot be assumed (as recognized by the commercial tests, too; see above). Similarly, languages other than English are unlikely to be found in a revised test; this step is perhaps too great given the test's history and, of course, its stated purpose.

### **Conclusion: A test for the American University ... of Beirut**

The paper has presented a reflection on the review of a language test that remained more or less static while its context of use evolved over an extended period. By discussing a rather unusual case, the review was able to show changes in perspective in the field of language testing over recent decades concerning, for example, construct definition and assumptions about test design. It highlighted the importance of being able to describe a test population and how context affects the appropriate uses of test scores. It also drew attention to the increasingly complex contexts in which university admission tests are employed.

The reflection may assist in the revision of the AUB-EN to serve effectively in its current context. It also indicates the scope for broader institutional innovation to support the test and clarify its purpose. Potential changes include more effective data collection about the language profiles of applicants and students, and the development of explicit institutional policies on language use. The outcomes of these projects are likely to be useful to other institutions dealing with similar issues, both English-medium universities in the region and US institutions enrolling greater numbers of students with diverse language backgrounds.

At AUB, the continued use of large-scale English for academic purposes tests (e.g., the TOEFL and IELTS) seems appropriate given the view that the university must preserve its comparability with US institutions – representing AUB as an *American university*. Nevertheless, a local English language proficiency test for admission is also desirable for practical reasons; this test, while aligned to the large-scale tests as closely as is feasible given available resources, should be designed to respond to local needs and reflect local practices – representing AUB as a university of *Beirut*.

## References

- Aintzane, D., Lasagabaster, D., & Sierra, J. M. (2013). *English-medium instruction at universities: Global challenges*. Bristol, UK: Multilingual Matters.
- Anderson, B. S. (2011). *The American University of Beirut: Arab nationalism and liberal education*. Austin, TX: University of Texas Press.
- Arnold, L. R. (2014). "The worst part of the dead past": Language attitudes, policies, and pedagogies at Syrian Protestant College, 1866–1902. *College Composition and Communication*, 66(2), 276–300.
- Baroudi, S. G. (1983). *Scores on the AUB EN and TOEFL as predictors of English grades at the AUB* (Unpublished MA thesis). American University of Beirut, Lebanon.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1–47.
- Chalhoub-Deville, M., Brown, A., & Willse, J. (2008, January). AUB admissions tests and placement program consulting report. (14 pp.). American University of Beirut, Lebanon.
- Chapelle, C. A., & Voss, E. (2014). Evaluation of language tests through validation research. In A. J. Kunnan (Ed.), *The companion to language assessment* (Vol. 3, pp. 1079–1097). Hoboken, NJ: Wiley-Blackwell.
- Davidson, F., & Cho, Y. (2001). Issues in EAP test development: What one institution and its history tell us. In J. Flowerdew & M. Peacock (Eds.), *Research perspectives on English for Academic Purposes* (pp. 286–297). Cambridge, UK: Cambridge University Press.
- Davies, A. (2013). *Native speakers and native users: Loss and gain*. Cambridge, UK: Cambridge University Press.
- Department of English. (2013, November 8). Faculty position in English language testing [Mailing list submission]. Message posted to LTEST-L@lists.psu.edu
- Dewey, M. (2012). Beyond labels and categories in English language teaching: Critical reflections on popular conceptualizations. In C. Leung & B. V. Street (Eds.), *English a changing medium for education* (pp. 129–149). Bristol, UK: Multilingual Matters.
- Educational Testing Service. (2007). *Test and score data summary for TOEFL Computer-Based and Paper-Based Tests. July 2005–June 2006 test data*. Princeton, NJ: ETS.

- Educational Testing Service. (2010, December). *Linking TOEFL iBT scores to IELTS scores – A research report*. (16 pp.). ETS, Princeton, NJ.
- Educational Testing Service. (2014). *A guide to understanding TOEFL iBT scores*. Princeton, NJ: ETS.
- Elder, C., & Davies, A. (2006). Assessing English as a lingua franca. *Annual Review of Applied Linguistics*, 26, 282–301.
- Elder, C., McNamara, T., Kim, H., Pill, J., & Sato, T. (forthcoming). Interrogating the construct of communicative competence in language assessment contexts: What the non-language specialist can tell us. *Language & Communication*.
- 'English Teaching Center in Beirut'. (1967). *The Linguistic Reporter*, 9(2), 2–3. Retrieved from <http://www.cal.org/resource-center/resource-archive/linguistic-reporter-collection>
- Facts and figures 2015* [Booklet]. (2014). Office of Communications [and] Office of Institutional Research and Assessment, American University of Beirut, Lebanon. Retrieved from <http://www.aub.edu.lb/main/about/Documents/facts-figures-15.pdf>
- Graduate Catalogue 2015–16*. (n.d.). Beirut, Lebanon: American University of Beirut. Retrieved from <http://www.aub.edu.lb/registrar/Pages/catalogue-graduate15-16.aspx>
- Haddad, M. D. (1995). *The validity and reliability of the AUB EN* (Unpublished MA thesis). American University of Beirut, Lebanon.
- Hanania, E., & Shikhani, M. (1986). Interrelationships among three tests of language proficiency: standardized ESL, cloze and writing. *TESOL Quarterly*, 20(1), 97–109.
- Harding, L. (2014). Communicative language testing: Current issues and future research. *Language Testing Quarterly*, 11(2), 186–197.
- 'History of IELTS' [Webpage]. (n.d.). International English Language Testing System. Retrieved from [http://www.ielts.org/researchers/history\\_of\\_ielts.aspx](http://www.ielts.org/researchers/history_of_ielts.aspx)
- Hulstijn, J. H. (2015). *Language proficiency in native and non-native speakers: Theory and research*. Amsterdam, The Netherlands: John Benjamins.
- Jenkins, J. (2011). Accommodating (to) ELF in the international university. *Journal of Pragmatics*, 43(4), 926–936.
- Jenkins, J., & Leung, C. (2014). English as a lingua franca. In A. J. Kunnan (Ed.), *The companion to language assessment* (Vol. 4, pp. 1607–1616). Hoboken, NJ: Wiley-Blackwell.

- Kreidler, C. J., & Pedtke, D. A. (Eds.). (1970, June). *Teaching English to speakers of other languages; United States activities: 1969* [Report]. (16 pp.). Washington, DC: Center for Applied Linguistics. Retrieved from ERIC database (ED040393).
- Lee, W. R. (1971). Ten years of the teaching of English as a Foreign Language. *ELT Journal*, 26(1), 3–13.
- McNamara, T. F. (1996). *Measuring second language performance*. London, UK: Longman.
- McNamara, T. (2014). 30 years on – evolution or revolution? *Language Assessment Quarterly*, 11(2), 226–232.
- 'New AUB-EN specifications' [Document]. (n.d.). Center for English Language Research and Teaching, American University of Beirut, Lebanon.
- Pill, J. (2015, May). The use of English language tests at the American University of Beirut [Report]. (21 pp.). American University of Beirut, Lebanon.
- Read, J. (2015). *Assessing English proficiency for university study*. Basingstoke, UK: Palgrave Macmillan.
- Sato, T. (2014). *Linguistic laypersons' perspective on second language oral communication ability* (Unpublished PhD thesis). University of Melbourne, Australia.
- Scott, M. S., & Tucker, G. R. (1974). Error analysis and English-language strategies of Arab students. *Language Learning*, 24(1), 69–97.
- Shaaban, K. A. (1997). Bilingual education in Lebanon. In J. Cummins & D. Corson (Eds.), *Encyclopedia of language and education, Volume 5: Bilingual education* (pp. 251–259). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Shikhani, M. E. (1982). *The cloze test as a possible supplement to the English Entrance Examination at the American University of Beirut* (Unpublished MA thesis). American University of Beirut, Lebanon.
- Shohamy, E. (2011). Assessing multilingual competencies: Adopting construct valid assessment policies. *The Modern Language Journal*, 95(3), 418–429.
- Spolsky, B. (1995). *Measured words: The development of objective language testing*. Oxford, UK: Oxford University Press.
- Stubbs, J. B., & Tucker, G. R. (1974). The cloze test as a measure of English proficiency. *The Modern Language Journal*, 58(5/6), 239–241.
- Sukhun, N. F. (1978). *The validity of selected variables in predicting success at the Faculty of Arts and Sciences of the American University of Beirut* (Unpublished MA thesis). American University of Beirut, Lebanon.

Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21–36.

*Undergraduate Catalogue 2015–16*. (n.d.). Beirut, Lebanon: American University of Beirut. Retrieved from <http://www.aub.edu.lb/registrar/Pages/catalogue-undergraduate15-16.aspx>

Xi, X., Bridgeman, B., & Wendler, C. (2014). Tests of English for academic purposes in university admissions. In A. J. Kunnan (Ed.), *The companion to language assessment* (Vol. 1, pp. 318–337). Hoboken, NJ: Wiley-Blackwell.