In the fall of 2009, the University of Calgary-Qatar (UCQ) undertook a research project funded by the Qatar Foundation’s Qatar National Research Fund (QNRF) Undergraduate Research Experience Program (UREP). The goal was to develop and assess the validity of an assessment rubric which could be used by institutions to assist in the selection of appropriate placement tests for EAP. This pilot project was the first stage of a longer project aimed at improving placement test procedures at UCQ.

At the time the project began, UCQ was using the paper-based Oxford Placement Test (OPT), which was also used on the main campus in Calgary, to place incoming students into remedial EAP classes if necessary. However, classroom teachers felt the test often did not place students accurately, particularly the listening section, and felt improvements in placement procedures were necessary. Although our student body was quite small at the time, we expected growth to make paper-based testing more challenging, and we therefore particularly wanted to explore computerized options.

Our research team included four undergraduate students; a primary goal of the project was to help them develop their research skills by actively participating in all stages of this project. Being new to the region, the faculty found student involvement in selecting a suitable test for the area to be invaluable.

Literature Review
The literature review was conducted collaboratively, simultaneously with the development of the rubric, by the faculty and student researchers in weekly meetings. One common theme arising from the literature review was a concern that a test taker’s culture might have an effect on the way test items are answered. Abu-Rabia (1998) found that Arab students learning Hebrew in Israel were more successful on reading tasks with content familiar to their Arab culture than on tasks with unfamiliar cultural content. In an interesting older study by Yousef (1968), a group of Arab students were able to understand the plot of literature, but misunderstood the motives of the characters due to their lack of familiarity with the culture.

Another issue arising from the literature review was the importance of a match between the program content with the content of the placement test. Brown (1989) found that a placement test used in one institution did not reflect learning in classes, and that the language ability of students promoted into a level did not resemble that of students placed into that level. In addition, tests which emphasize English used in business settings or English necessary to immigrants moving to an English-speaking country may result in placements not appropriate to the academic EFL context of UCQ classrooms.

Weighting of test sections was also an area of interest during our literature review. Most placement tests include a grammar section which is usually relatively heavily weighted. However, Santos (1988) found that higher levels of grammatical inaccuracy were tolerated in the writing and speech of Non-native Speaking (NNS) students than of Native Speaking (NS) students. Thus, in contexts such as ours, where the majority of students use English as an additional language, instructors may be less demanding of grammatical accuracy, creating a mismatch between the language being tested on placement tests and the language necessary for success in classrooms.
Methodology
The research was conducted in four stages. First, the literature was reviewed, and a placement test selection rubric was developed. The Placement Test Selection Rubric was then used to review six placement tests. Next, a number of student volunteers took one test which scored well on the rubric, and two which scored poorly. Finally, a group of teachers familiar with the students evaluated the students’ language level according to their familiarity with the students’ classroom work. If the rubric was functioning well, we theorized that the highest scoring placement test would give the best agreement with the teachers’ evaluations.

Stage 1: Rubric Development
The literature review and rubric development were conducted simultaneously using an iterative process over several months. Each team member reviewed a number of articles, and at weekly meetings we reflected on key points arising from the articles, as well as our program needs and the characteristics of our student body. Using a number of idea-generation and organization strategies such as brainstorming and concept mapping, themes emerging from the discussions were organized and weighted into what became the final sections of the Placement Test Selection Rubric and the criteria within each.

The final rubric addressed issues of cultural appropriateness, test instructions and administration, reliability and validity, and test range (reading, writing, listening, speaking, grammar and vocabulary). Sections were weighted such that the emphasis was on test range, at 50% of the final rubric score, and cultural appropriateness at 25%. The rubric was created as an Excel document with formulas to compute the scores for each section.

To help ensure validity, a second group of language testing experts was asked to review the rubric and provide feedback. This resulted in the inclusion of qualitative elements, comments boxes for each section, as well as a separate Placement Test Summary Sheet which were added to allow users of the rubric to include more detailed comments when applying the rubric.

Stage 2: Placement Test Review and Selection
Using the rubric, six placement tests were reviewed by the research team (see Figure 1 below). As we did the computerized tests together as a group, the student researchers discussed each test item. The faculty researchers were thus able to see which questions and sections posed problems for them. Rubric scores were gathered and averaged for each test. We also gave each test an intuitive holistic score based on the qualitative comments from the comments boxes and Placement Test Summary Sheet as a way of checking whether the rubric score matched the feelings of the research team as to which test was most suitable.

The highest scoring placement test was selected for trials with students as well as two of the lowest scoring placement tests. This resulted in the selection of one computer-based test with the highest score and two tests (one computer and one paper-based) with the lowest scores. Ideally, two high-scoring and two low-scoring tests would have been selected, but given our limited placement test pool and student test subject population, we felt three tests would be sufficient for our purposes.

Stage 3: Deployment of Placement Tests
Students from the general student population at the UCQ were recruited to complete all three tests in semi-random order.

Stage 4: Expert Raters
The language level of the student participants was evaluated by 11 expert raters, who were teachers familiar with the students’ language levels. The raters were asked to place students into one of five categories based on the UCQ English for Academic Purposes (EAP) program levels.

For each test, comparison groups were created by placing cut-scores for each level at the point that gave the best possible agreement with the expert raters. The paper-based OPT was in use at UCQ at the time and the cut-scores created for this test using this method were almost identical to the cut-scores in use. The resulting level placements according to each test were then compared to the level placements made by the expert raters using a Kappa test. For the purposes of placement, it was decided that a fairly high weighted Kappa score of .70 as a measure of agreement between expert raters and the placement test was desirable.
Results

Test Reviews

Accuplacer

Accuplacer ESL is an American computer adaptive test produced by the College Board and is widely used by colleges and universities throughout the United States. It consists of five ESL sections: Reading Skills, Sentence Meaning, Language Use, Listening, and the WritePlacer ESL. Accuplacer is an Internet-based test, which means that no software must be downloaded and IT requirements are minimal.

While we encountered a number of Americanisms (Miranda Rights, American city names), our student researchers felt that their ability to answer the questions was not compromised due to these elements. The multiple-choice format used throughout was also familiar to our students.

Both examples and video tutorials were provided to help test takers. The examples for each section were clear and simple. However, the video tutorials were very unhelpful. The screenshots were taken from a Mac computer, and as we did the test on a PC, the images did not resemble what we saw at all. The tutorials also contained extremely complex language. For example, a tutorial on how to use the computer’s mouse instructed that one’s “index finger rests on the primary button” and one on using the keyboard instructed users to “use the modifier keys on your keyboard to perform certain actions.” These instructions would be beyond the linguistic level of many of our incoming students.

In our trial, the test seemed appropriate in academic content for our program. The questions in the Reading Skills section were based on readings of, at most, a few sentences, but tested a variety of skills we practice in our EAP program including making inferences, distinguishing fact and opinion, and interpreting metaphor; it included fictional as well as factual texts. Items in the listening section were based on short, realistic conversations in a variety of settings. Test takers could play both the dialogue and question three times. Pictures gave the conversations a bit of context. The Language Use and Sentence Meaning sections seemed to be fairly standard tests of grammar and vocabulary. We would have liked to have seen more items from the Academic Word List in the Sentence Meaning section, but perhaps their absence was a result of the computer-adaptive nature of the test and the way we answered the initial questions, as we did not attempt to answer all of the questions correctly.

The computerized writing test includes a short reading passage for context before the prompt. Helpful features include a description for test takers as to what elements of their writing will be graded, a timer that shows the amount of time left, and an icon that can be clicked on to show word count. The prompts we saw in our trials were challenging enough that our lower level students might not have been able to respond well to them. We would have liked to see some easier prompts offered to better discriminate among the lower level students we sometimes work with.

Compass

Compass is also an American computer adaptive test and was developed by the testing organization ACT. We found the realistic academic content of the test to be perhaps slightly more suitable than that of the Accuplacer but the IT requirements of the test were troublesome enough to make the Accuplacer preferable overall, as described below.

Compass has four sections: Listening, Reading, Grammar/Usage, and Essay. The listening items we encountered in our trial were based on natural-sounding dialogues that progressed as items were answered correctly from very short, simple conversations in familiar contexts to rather long abstract dialogues that tested skills like making inferences and recognizing metaphor in texts that resembled those that might be encountered in a range of first year university courses. Reading texts similarly progressed from very short texts to longer academic texts with questions testing skills such as finding main ideas and details, making inferences, and drawing conclusions. The grammar test included questions based not only on sentence-level grammar, but also on the relationship between ideas in paragraphs. The computerized writing test seemed to provide satisfactory prompts and an adequate time limit.

The Compass testing package runs on a secure browser that prevents test takers from accessing the internet while doing the test. Unlike Accuplacer, Compass requires that its software be downloaded.
by an institution’s IT department onto each computer used to run the test. Running the tests also requires reconfiguring the screen resolution for each computer, a process that is simply not practical in a busy, shared computer lab. For a school with a dedicated testing lab, this may be less of a concern, but these technical issues ultimately resulted in the decision that this test would not work for us.

**Password**

Password, developed by English Language Testing, Ltd., consists of five sections plus an optional online writing exam. It is not computer adaptive, although the items for each test are drawn from a data bank so that each test is unique. Apart from the writing section, it measures only grammatical knowledge and vocabulary with the rationale, according to Password promotional materials, that these sections of tests are usually most reliable (Accessible, 2011). Password was designed particularly for students who may have studied English extensively, but not have had many opportunities to practice in daily life or in skills-based classrooms. The test developers feel that this type of student may be unfairly disadvantaged by communicative-style tests, particularly in listening and speaking, but that communicative skills develop quickly in an English-speaking environment (Accessible, 2011).

This profile does not match our student population well. As many Gulf Arabs need to use English with expatriate workers in their daily lives, their communicative abilities are often quite advanced. However, they may be disadvantaged by tests of formal grammar.

In addition, our student researchers found this test quite stressful. One test section deals entirely with collocations. Despite having exceptionally strong English, our student researchers did surprisingly poorly on this section. They knew the meaning of all the words presented, but were simply unable to pick out the most common word partners. We theorized that this might have been due to their using English primarily as a lingua franca and having less exposure to the language as spoken by native speakers. This section was very demoralizing to our student researchers.

The fifth section of the test requires test takers to pick which sentences out of a set of four or five are grammatically correct. Our student researchers found this section confusing and exhausting. They were unsure whether they could pick more than one sentence, and the amount of reading required at this stage of the test was quite tiring.

In addition, the method of answering questions was unfamiliar and a little confusing to the team. Test takers must click on numbers at the bottom to choose which question they want to work on. Test takers can flag questions they want to return to, although it took us a while to figure this out.

**Oxford Online Placement Test**

Oxford Online Placement Test is a computer-adaptive placement test by Oxford University Press. It has only two sections. The Use of English section covers grammar, vocabulary, and reading. The Listening section covers listening skills.

This test was probably the easiest to conduct. The test can be done on any computer: test takers simply go to the website and enter a password to begin. A drawback to the test’s ease of administration is a lack of control while test takers do the test. The test is not delivered in a secure browser, so test takers can access the internet while doing the test. During our trials of the test, despite having several proctors in the room, we found that several students accessed Google Translate.

The focus of this test was less academic than the other three, with test items on general English for daily life and business as well. Although the listening dialogues were realistic, they covered a wide range of situations that our students would likely not need to deal with. We felt this test would be more suitable for students moving abroad to study in an English speaking country.

The test website claims the test is designed to test how learners use their knowledge of grammar and vocabulary to interpret the meaning of English for real communication. Our team, however, found that the test had a very heavy emphasis on idioms, such as “throw my hat in the ring” and “call it a day.” Some idioms even stumped the faculty researchers, for example, “turn up for the books.” Perhaps these were specific to British English. As our students study in classrooms where most of their classmates also use English as a second language, familiarity with these idioms is less important for them than for students studying in English speaking countries.
Other British expressions such as “straightaway,” “got sacked,” and “shared a flat” were confusing for our student researchers as well, as they were less familiar with British English. The question about sharing a flat was difficult for the student researchers on another level as well, as they were not able to interpret the relationship between the speakers, being unfamiliar with this living arrangement. British accents added some difficulty as well, and overall the student researchers found this test rather demoralizing.

University of Michigan English Placement Test

The University of Michigan English Placement Test is a paper-based test which includes sections for listening, reading, grammar, and vocabulary. In the listening section, test takers hear a statement and must choose the most appropriate response or best paraphrase from options printed in the test booklet. Although the format of the listening section seemed straightforward to the faculty researchers, the student researchers found it quite confusing, and in fact asked that the test be stopped and several questions demonstrated.

Reading questions are in multiple-choice format and based on short passages of at most a few sentences. As the passages are so short, skills like identifying relationships between paragraphs cannot be measured. The grammar and vocabulary sections are also in multiple-choice format. Words from the Academic Word List are common.

The instructions for all sections of the test are together on the same page. Our student researchers commented that they would have preferred to have the instructions for each section printed above each section.

Oxford Placement Test (OPT)

The listening section of the paper-based OPT is composed of 100 questions and requires test takers to choose which of a pair of similar sounding words they hear, for example the difference between “oarsman” and “horseman.” Some vocabulary in the listening section is so low frequency that even the most advanced students at UCQ would not be familiar with it, for example “timberworks,” and “barrel-load.” This in itself would not necessarily invalidate the test, but our student researchers found it very distracting. The faculty and student researchers alike expressed some doubts as to whether the ability to distinguish words differing by only a phoneme could adequately predict the complex listening skills required to function at UCQ. Both faculty and student researchers also found it very difficult to maintain concentration for the entire ten minutes required, and each of us missed a number of items.

The grammar section also contains 100 questions, and like the Michigan English Placement Test is delivered in multiple-choice format. The items focused on typical verb tense and sentence structure points taught in most ESL classes.

Data from Placement Test Ratings

Using the Assessment Rubric

Using the Placement Test Selection Rubric, the research team assigned scores to the six placement tests reviewed, as in Figure 1. The score was based on the rubric, and used to choose the tests to pilot. The holistic score was based on the Placement Test Summary Sheet and served as a confirmation that the rubric matched the research team’s opinions as to the best test.

<table>
<thead>
<tr>
<th>Test Name</th>
<th>Test Mode</th>
<th>Rubric Score /100</th>
<th>Holistic Score /10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuplacer</td>
<td>Computer</td>
<td>89</td>
<td>8</td>
</tr>
<tr>
<td>Compass</td>
<td>Computer</td>
<td>85</td>
<td>9</td>
</tr>
<tr>
<td>University of Michigan English Placement Test</td>
<td>Paper</td>
<td>71</td>
<td>8</td>
</tr>
<tr>
<td>Password</td>
<td>Computer</td>
<td>66</td>
<td>7</td>
</tr>
<tr>
<td>OPT Online Test</td>
<td>Computer</td>
<td>51</td>
<td>5</td>
</tr>
<tr>
<td>OPT Placement Test</td>
<td>Paper</td>
<td>46</td>
<td>5</td>
</tr>
</tbody>
</table>
Deployment of Placement Tests
The tests selected for piloting were the top-scoring test, Accuplacer, and the two lowest-scoring tests, Oxford Online Placement Test and the paper-based OPT. Accuplacer was completed by 16 students, Oxford Online Placement Test by 26 students and OPT by 23 students.

Correlations
We were interested in determining whether any sections of the tests we tried were strongly correlated to sections on other tests (see Appendix). The most striking finding was that the correlation between the paper-based OPT Listening section and every section of every other test was close to zero. Even with small numbers of students taking the tests, this made us suspect that the listening section of the OPT was functioning randomly with our student population, supporting our earlier feelings that it was not a valid test of listening skills for our program.

<table>
<thead>
<tr>
<th></th>
<th>Accuplacer</th>
<th>OPT Online</th>
<th>OPT Paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohen's Kappa</td>
<td>.3766</td>
<td>.2632</td>
<td>.2477</td>
</tr>
<tr>
<td>Weighted Kappa (linear)</td>
<td>.4839</td>
<td>.4257</td>
<td>.2562</td>
</tr>
</tbody>
</table>

Figure 2. Agreement between expert raters and placement test scores.

Typing Speed
One aspect of computerized testing to which we did not give enough weight while reviewing tests was the effect of the poor typing ability of incoming students. One faculty researcher gave a typing speed test to a group of fairly new students at the institution after our test trials were underway, and found an average typing speed of only eleven words per minute, probably low enough to have a serious effect on computerized writing test scores.

Discussion
The purpose of this study was to determine whether a rubric could be designed to effectively indicate the most suitable placement test for use at UCQ. Accuplacer was the highest scoring test according to the rubric, and OPT Online Test and OPT Placement Test were lowest. If the rubric was effective at choosing the best test, it was predicted that its placements would show the highest level of agreement with the expert rater placements, while the level of agreement for the OPT Online test and the paper-based OPT would show the least.

While the Accuplacer did achieve the highest agreement with the expert raters, none of the tests had a high enough level of agreement to validate the Placement Test Selection Rubric as a tool for placement test selection, or to indicate any of the tests as acceptable for use at our institution. Higher numbers of students taking each test would be required to know whether this poor agreement was due to poor functioning of the rubric or other factors, such as a difference in the language skills measured by the tests and the language skills the expert raters deemed necessary for success at UCQ.

The most valuable information in terms of the institution came from working through samples of each test with our four student researchers. Their feedback and thoughts while working through each test sample provided invaluable insights into how
our student population would likely approach each test. For example, faculty was able to observe that even for very high level students, questions based on idioms and collocations were exceptionally difficult. It was theorized that as English is used primarily as an additional language by most of the student population, they have less exposure to these aspects of English and may have less need for them in their classrooms.

Conclusion
Although this was a small study with limited numbers of participants, our results indicate a need for more research into the question of whether students studying in an EFL context perhaps require a slightly different set of language skills than those studying in classrooms dominated by native speakers. The poor performance of our research students on items related to collocations and idioms, despite their high language level and success as students at UCQ, suggests a need for research into whether these aspects of language are less necessary for success in EFL settings.

It was primarily on the basis of the results of this study that Accuplacer has been adopted as the placement test at UCQ and is in the process of being implemented. An on-going study is currently in place to set appropriate entry, exit and EAP level cut scores.

For more information, please contact Virginia Christopher at vlchrist@ucalgary.ca, Brad Johnson at bjohnso@ucalgary.ca, or Karen Brooke at kbrooke@vcc.ca. To obtain a copy of the Placement Test Selection Rubric and the Placement Test Summary Sheet, please contact Karen Brook at kbrooke@vcc.ca.

References


Appendix
Results: Correlation Analyses

<table>
<thead>
<tr>
<th></th>
<th>OPT (paper)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Listening</td>
<td>Grammar</td>
<td>Total</td>
</tr>
<tr>
<td>Accuplacer</td>
<td>0.07992</td>
<td>0.69098</td>
<td>0.62999298</td>
</tr>
<tr>
<td>Language Use</td>
<td>-0.154</td>
<td>0.74158</td>
<td>0.54316763</td>
</tr>
<tr>
<td>Listening</td>
<td>-0.109</td>
<td>0.66164</td>
<td>0.5003664</td>
</tr>
<tr>
<td>Reading Skills</td>
<td>0.18287</td>
<td>0.73404</td>
<td>0.72358122</td>
</tr>
<tr>
<td>Sentence Meaning</td>
<td>0.10428</td>
<td>0.65195</td>
<td>0.61041479</td>
</tr>
<tr>
<td>Writeplacer</td>
<td>0.01529</td>
<td>0.79029</td>
<td>0.67832386</td>
</tr>
<tr>
<td>Total Excluding Writeplacer</td>
<td>0.01529</td>
<td>0.79029</td>
<td>0.67832386</td>
</tr>
</tbody>
</table>
20 students did both the Oxford Placement Test (OPT) and the Oxford Online Placement Test.

<table>
<thead>
<tr>
<th>OPT (paper)</th>
<th>Oxford Online Placement Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Listening</td>
</tr>
<tr>
<td>Listening</td>
<td>-0.0272</td>
</tr>
<tr>
<td>Grammar</td>
<td>0.76837</td>
</tr>
<tr>
<td>Total</td>
<td>0.63283</td>
</tr>
</tbody>
</table>

15 students did both the Oxford Online Placement Test and the Accuplacer.

<table>
<thead>
<tr>
<th>Accuplacer</th>
<th>Oxford Online Placement Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Listening</td>
</tr>
<tr>
<td>Language Use</td>
<td>0.45369903</td>
</tr>
<tr>
<td>Listening</td>
<td>0.62745738</td>
</tr>
<tr>
<td>Reading Skills</td>
<td>0.7079608</td>
</tr>
<tr>
<td>Sentence Meaning</td>
<td>0.49654448</td>
</tr>
<tr>
<td>Writeplacer</td>
<td>0.44179812</td>
</tr>
<tr>
<td>Total Excluding Writeplacer</td>
<td>0.63601477</td>
</tr>
</tbody>
</table>

Karen Brook is an ESP instructor with Vancouver Community College. She worked at the University of Calgary-Qatar for two years and is interested in assessment.

Mona Ali Aden is a nursing student at University of Calgary-Qatar, an executive member of the Nursing Student Society-Qatar, and involved in several research projects.

Noof Al-Kuwari is a nursing student at University of Calgary-Qatar, president of the Nursing Student Society-Qatar, and student representative to the UCQ research committee.

Virginia Christopher instructs in the University of Calgary-Qatar EAP program. She has worked as a high-school teacher, ESL/EFL instructor, program manager and teacher educator.

Mihad Ibrahim is a nursing student at University of Calgary-Qatar and president of the debate club at UCQ.

Dr Brad Johnson is Director of the Faculty Development Centre at University of Calgary-Qatar and works with faculty to integrate innovative pedagogy and technology into practice.

Oumaima Souyah is a nursing student at University of Calgary-Qatar. She loves volunteering in her field and university and hopes to pursue a doctorate.

❉❉❉❉❉