A Case for Faculty Involvement in EAP Placement Testing

Cindy James and Elizabeth Templeman

The EAP placement procedure at Thompson Rivers University (TRU) involves multiple measures to assess the language skills of incoming students, some of which are facilitated and all of which are assessed by ESL faculty. In order to determine the effectiveness of this comprehensive EAP placement process and the effect of the faculty factor, a placement validity study was conducted at TRU. The results revealed that with faculty involvement, the accuracy of course placement ranged from 81% to 84; without it the accuracy of the placement would have decreased to between 44% and 66%, depending on the skill area.

À la Thompson Rivers University (TRU), la procédure de classement pour les cours d'anglais à des fins académiques (EFA) implique plusieurs mesures pour évaluer les aptitudes langagières des nouveaux étudiants. Certaines de ces mesures sont facilitées; toutes sont évaluées par les professeurs en anglais langue seconde. Afin de déterminer l'efficacité de cette procédure compréhensive de classement et l'influence de l'implication du personnel enseignant, une étude de la validité des placements a été entreprise à TRU. Les résultats indiquent qu'avec l'implication du personnel enseignant, la fiabilité diagnostique du placement dans les cours variait de 81% à 84%; sans cette implication, la fiabilité aurait chuté à entre 44% et 66%, selon les compétences évaluées.

Introduction

Language proficiency assessment for international applicants is common practice for most educational institutes and usually involves administering some type of commercially available or locally developed English placement examination (Alderson, Krahnke, & Stansfield, 1987; Chalhoub-Deville & Turner, 2000; Kahn, Butler, Weigle, & Sato, 1994; Paltridge, 1992; Person, 2002; Rees, 1999; Roemer, 2002; Seaman & Hayward, 2000). If the language assessment indicates that the applicant does not have the necessary skills, many institutes offer English for Academic Purposes (EAP) programs to develop these skills. In this situation, the purpose of the language proficiency testing shifts from admission to placement. Although placement testing may not be as high-stakes as admissions testing, its effect is significant. For students it can delay or prolong their studies; for EAP programs inappropriately placed students may jeopardize the quality and integrity of the course sequence. Therefore, accurate EAP placement testing is crucial. To date, most

of the empirical research has focused on the predictive validity of language proficiency testing for admissions rather than for EAP placement (Ayers & Quattlebaum, 1992; Light, Xu, & Mossop, 1987; Johnson, 1988). One exception was a study conducted by Wall, Clapham, and Alderson (1994) in which they explored the concurrent validity of the Lancaster University Institute for English Language Education (IELE) placement test. This study found significant correlations between students' self-assessments and their scores on the IELE reading writing and listening subtests (r=.30, .41, .51 respectively, p<.05); and between language teachers' assessments and all the IELE subtests (r range from .47 to .78, p<.01). However, Wall et al. concluded that little could be inferred from these correlations and that further research into validating placement testing was needed. Another study of California's English as a Second Language (ESL) Placement procedures by Kahn et al. (1994) revealed that most adult education agencies in California used a commercial or agency-made test to place ESL students and that over half these agencies were satisfied with the accuracy of their placement. This study did not provide any correlation data between test scores and achievement in the ESL programs/courses offered by the agencies.

Although predictive validity research on language placement testing is definitely lacking, there appears to be some consensus among researchers about methodologies for language proficiency testing and assessment. Fulcher (1999), for example, stated that EAP testing should focus on construct validity rather than test content and be scored according to the students' general abilities. Clapham (2000) concurred, but also suggested that EAP testing should include specific examples of English grammar and tasks such as essay-writing. Moreover, she recommended that EAP proficiency tests be replaced by aptitude tests, because "such tests would show how capable students were of swiftly absorbing and producing the academic discourse styles required in Western tertiary institutions" (p. 519). Meanwhile, Wall et al. (1994) recommended a post-admissions screening system in which the language proficiency of international students would be assessed on arrival. Spolsky (1997) supported multiple testing and alternate assessment methods and emphasized the importance of interpretation. Rees (1999) concurred and recommended a "multi-judgment approach to the assessment of international students' language proficiency" (p. 436), which would include such measures as aptitude information, instructors' assessments, subject-related assessments, and self-assessments.

As for the use of technology—specifically computer-based testing—in EAP placement testing, the published information appears to be limited. However, there is a profusion of research on the use of technology for assessing general language skills, most of which involves reviews, critiques, and/or comparisons of the various computerized language tests, including some of the most prominent standardized test such as the computer-based

Test of English as a Foreign Language (TOEFL-cbt) and its successor, the Web-based TOEFL_iBT, the International English Language Testing System (IELTS), and the Test of English for International Communication (TOEIC, Bachman, 2000; Brown, 1997; Chalhoub-Deville & Deville, 1999; Chapelle, Jamieson, & Hegelheimer, 2003; Dunkel, 1991, 1999; Sawaki, 2001; Stricker, 2002; Taylor, Jamieson, Eignor, & Kirsch, 1998).

Although the discussions and conclusions from these studies vary, there is consensus on some issues such as the recognition of the potential advantages and disadvantages of computerized language testing. Enhanced security, flexible testing arrangements, greater standardization, immediate feedback, more accurate scoring, multimedia presentation, and electronic data storage are some of the most notable advantages. The ability to assess a wide range of skill levels is another positive attribute of computerized language tests including computer-adaptive technology, because this type of testing system automatically adjusts the difficulty of the test to the skills of the individual examinee. The recognized drawbacks generally include complications involved with administering tests via computers (cost of infrastructure, access to computers and the Internet, etc.), as well as the potential negative effect on students who may not be familiar or comfortable with this technology. Given the proliferation of computerized language testing, one other area of agreement is that no matter what its purpose may be—instructional aid, admissions or placement test, pre- and post-testing—these tests need to be validated by both the producers and the users.

At Thompson Rivers University (TRU), the English as a Second Language (ESL) department incorporates in its EAP placement process for international students many of the features recommended for language testing in the context of a computer-based testing system. Specifically, TRU includes a post-admissions assessment that involves multiple measures: an oral interview and all five subtests from the Web-based *ACCUPLACER ESL* testing system. It also relies on participation by ESL faculty for conducting the oral interview, independently assessing the writing sample, and interpreting the placement scores. In order to determine the efficacy of this comprehensive EAP placement process, and specifically the effect of the *faculty factor*—as well as contributing to the limited research—a placement validity study was conducted at TRU. For this study the two key research questions were as follows:

1. How accurate is the EAP placement process at TRU?
2. To what extent does the accuracy of the process depend on faculty involvement?

Methods

This study took place at Thompson Rivers University (TRU) in Kamloops, British Columbia: a public educational institute offering a variety of univer-

sity, college, and technical programs including a comprehensive preparatory English as a Second or Additional Language program (ESAL). Two hundred, ten ESAL students from three intakes (winter, summer, and fall) in 2004 participated in this study. The average age of these students was 22.0 (*sd*=5.63) and there was a fairly even distribution of males (56%) to females (44%). The primary data sources were students' scores on the five *AC-CUPLACER ESL* tests and the oral interview and their final grades in their first semester of ESAL courses.

Program and Course Descriptions

The ESAL program at TRU is designed to provide specific and appropriate language instruction for EAP learners who intend to proceed to postsecondary study. Successful completion of the program means that a student has a sufficient level of English language proficiency to undertake studies at English speaking colleges or universities, and as such provides students with direct entry into other programs at TRU. The ESAL program has five levels. At the time of this study, the first level of reading and writing were integrated in one course, and the speaking, listening, and grammar were also integrated in a single communication course. At the next level, the reading and writing streams separated. At the highest three levels, the skill components were all separated. In addition, students at Levels IV and V were permitted to take one or two courses outside of EAP, including university or career-technical courses, providing that additional prerequisites for these courses had been met.

Students are placed in the ESAL program according to their individual skills, and their overall skill profile. They are considered to be in the level where most of their scores fall, but can be placed in an individual course that is either one level above or below that level. Thus a student may have four courses at Level III and also be taking a Level II or Level IV reading course, for example. However, program regulations do not allow a student's course load to have more than a one-level split. The reasoning behind this is that although individual courses focus on separate skills given the nature of language study, no discrete skill is studied and demonstrated in isolation. For example, reading comprehension is demonstrated in speaking and writing. Writing is demonstrated, particularly at higher levels of EAP, through responding to a written prompt, which requires competence in reading. Certain skills such as summary and paraphrase bridge skills and depend on a confluence of abilities.

The Assessment Tool and Procedures

The *ACCUPLACER ESL* testing system is a computer-adaptive, Web-based program marketed by the College Board (2007). These tests assess the English skills of students who have learned English as a second language or who are

native English-speakers with limited proficiency. It consists of the following five tests, all of which are used at TRU and were included in this study: *Reading Skills* evaluates comprehension of short passages; *Language Use* measures grammar and usage; *Sentence Meaning* assesses the understanding of word meanings in one- or two-sentence contexts; *Listening* measures the ability to listen to and understand one or more speakers; and *WritePlacer ESL* provides a direct measure of writing skills. Except for *WritePlacer ESL*, all the tests consist of 20 multiple-choice questions that are scored on a scale of 0 to 120. The *WritePlacer ESL* test requires the examinee to complete a writing sample based on a randomly assigned prompt and is electronically scored by *IntelliMetric,* a product of Vantage Learning. Using a holistic scoring process designed to emulate human scorers, this system rates each essay on a scale of 0 to 6 based on its overall effectiveness (Elliot, 2003). Detailed descriptions of each test are provided in Appendix A. The scoring rubric for the *WritePlacer ESL* test is provided in Appendix B.

The *ACCUPLACER ESL* testing system was adopted by TRU for several reasons: it was fairly straightforward to administer; the results were available immediately; it tested all relevant skill areas; it was capable of assessing a wide range of skills; and it was relatively inexpensive.

The *ACCUPLACER ESL* tests were administered to the participants in the morning on the first day of their orientation week for each intake. Some students completed the tests in 90 minutes; most took approximately two hours; and a few took three full hours. That same afternoon, small-group oral interviews were conducted, and the compositions were independently assessed by the faculty using the *ACCUPLACER* scoring rubric (Appendix B). After the lunch break, students were directed to their oral interviews. A copy of their compositions and a report of their raw scores were sent from the Assessment Centre to the ESL Department Chair.

Interview

For the interview, students with similar scores on the *Language Use* and *Listening* tests were sorted into small groups of three to five. Each of these groups was interviewed for approximately 15 minutes by two ESL faculty members, one facilitating and one observing. The facilitator asked open-ended questions and provided conversation prompts appropriate to the level of the students. Questions such as "Where are you from?" or "How was your trip?" were typically used for lower-level students, whereas questions such as "What do you think is a serious problem facing your country today?" or prompts such as "Tell us about your goals" were used for higher-level students.

During the interview, the observer recorded comments on a range of criteria including fluency, pronunciation, syntax, vocabulary, and level of interaction. Immediately following each interview, the faculty summarized

their comments and observations and assigned a single oral score to each student using an oral interview scoring rubric (Appendix C).

Placement Procedures

For this study standard TRU EAP placement practices were adhered to. The ESL faculty followed a multistep process, placing students by their discrete skill scores into the appropriate level for each skill. Initial placements into the writing courses were based on the writing samples from the *WritePlacer ESL.* Working in small groups, faculty independently assessed the writing samples using the same scoring rubric (Appendix B) before comparing their scores with those of *ACCUPLACER WritePlacer ESL.* Matches were accepted; discrepancies were resolved by an additional faculty reader. This additional reader, using both scores and the associated comments, circled and initialed the mark she or he agreed with, recording comments to support the decision. If there was no agreement or considerable discrepancy, the markers conferred with an additional marker and arrived at a compromise.

Initial placement into the communication course(s) was based on scores from the *Language Use* and *Listening* tests and the oral interview score. Initial placement into the reading courses was determined by combined scores from the *Reading Skills* and *Sentence Meaning* tests.

Ultimate student placement into EAP courses was determined by a team of instructors who reviewed all assessment results—*ACCUPLACER ESL* subtest scores, oral scores, adjusted composition scores, and recorded comments—on a student-by-student basis. In many cases, the initial placements were accepted and no further scrutiny was required. However, numerous students in this pilot had scores scattered over four levels; hence their placements had to be reevaluated in accordance with program policy restricting students to courses in no more than two levels.

In such cases faculty would consider how related factors—such as whether students' listening skills would facilitate comprehension at a higher level of syntax, or their writing skills were sufficient to meet the writing requirements of a given reading course—might affect target course placements. For example, if a student placed in level one writing and level four speaking—with grammar and reading at level three—he or she would perhaps be placed in level two writing, but held back to level three in speaking, with the remaining courses at level three. If this same student had placed in level two grammar and reading, the decision would most probably have been to place the student in levels one and two courses. The variety of factors was broad and the decisions complex, but with an intimate awareness of the program and with a comprehensive set of assessment data, faculty groups worked through each case fairly efficiently, each small group placing up to 30 students in an hour. Highly unusual placements were noted, with concerns recorded in the student's file to share with relevant instructors.

Data Analysis

To assess the distribution of scores, descriptive statistics were computed for each of the five *ACCUPLACER ESL* tests as well as the oral interview and the faculty essay scores. To assess the correlation between the computer-generated essay scores using *IntelliMetric* and the faculty-generated essay scores, regression analysis was employed and the level of agreement was calculated. The level of agreement refers to the percentage of identical essay scores assigned by an expert human scorer and the automated scoring system (exact agreement rate), or within one point of one another (adjacent agreement rate). Because this study focused on whether students would succeed based on their placement—not the degree to which they performed—final letter grades were classified as signifying success (C+ or better in the course), or nonsuccess (C or lower, Withdrew, or Did Not Complete). This dichotomous response variable was cross-tabulated with the placement by course group (reading, writing, and communication) to determine the effectiveness of the placement process. Each placement was classified as a match if the student took the course he or she was placed into based on the faculty assessment and a non-match if the student took a course other than that in which he or she was placed. The latter situations occurred when students successfully challenged their placement, usually by using an alternate test score such as TOEFL or IELTS. The accuracy of the placement, reported as a percentage, was calculated three ways: one for how well it correctly predicted success, another for how well it correctly predicted nonsuccess, and a third to rate the overall success of the placement procedure. Placements were considered correct when students took the courses recommended and were successful (i.e., the accuracy of predicting success), or when students took courses not recommended and were unsuccessful (i.e., the accuracy of predicting nonsuccess). Placements were considered incorrect when students took courses not recommended and succeeded or took courses recommended but did not succeed.

To assess the effect of faculty involvement, the response variables for the same course groupings were then cross-tabulated with the predicted placements based solely on the *ACCUPLACER ESL* cut scores without faculty interpretation or adjustment. In this case, each placement was considered a match if the student took the course in which he or she would have been placed based on one or more of the test score(s)—no faculty involvement—and a non-match if the student took a course other than that in which he or she would have been placed based solely on one or more of the test score(s). Again, the success, nonsuccess, and overall accuracy as percentages by course were reported to facilitate comparisons.

CINDY JAMES and ELIZABETH TEMPLEMAN

Results

Descriptive Statistics

The descriptive statistics for the tests are provided in Table 1. The distribution of the *ACCUPLACER ESL Listening* test, the faculty scores of the writing sample, and the oral interview scores were approximately normal, whereas the *ACCUPLACER ESL Language Usage, Reading Skills, Sentence Meaning,* and *WritePlacer ESL* test scores were negatively skewed, and the *ACCUPLACER WritePlacer ESL* test scores were non-normal.

Correlational Data for the Writing Samples

In this study a significant positive correlation was found between essay scores generated by faculty and those generated by the computer (*r*=.706, *p*<.001). In comparing the individual values generated by the faculty scorers with the computer scores, 39.5% were an exact match and 88.1% were adjacent (within one point or less of each other). Only 11.9% of the scores differed by two points or more.

Placement

The 210 students who participated in this study enrolled in a total of 580 ESL courses. Most enrolled in one reading course, one writing course, and one or two communication courses (integrated communication for the lower levels and separate grammar and oral communication for the upper levels courses).

Placement into reading courses. Of the 210 participants, 164 were placed in an ESAL reading course, and the rest were exempted based on *AC-CUPLACER ESL* or other test scores falling beyond the EAP range. As indicated in Table 2, with faculty interpretation of the *ACCUPLACER ESL*

Table 1

Descriptive Statistics of the Placement Test Scores

| | | ACCUPLACER ESL Tests | | | | Faculty Assessments | |
Statistic	Listening	Language Usage	Reading Skills	Sentence Meaning	Write Placer ESL	Essay	Oral Interview
Mean	70.1	87.0	88.4	83.0	3.4	3.0	2.9
Standard deviation	17.27	19.21	21.39	21.34	1.10	1.20	1.20
Median	69.5	90.5	90.5	87.0	3.0	3.0	3.0
Minimum	20.0	25.0	26.0	25.0	0	0	1
Maximum	119.0	117.0	119.0	119.0	5.0	6.0	5.0
Max. score possible	120	120	120	120	6	6	5

Reading Skills and *Sentence Meaning* scores, 136 of the 141 successful students were placed correctly (96.5% correct); however, only two of the 23 nonsuccessful students were unsuccessful as predicted (8.7% correct), resulting in an overall course placement accuracy of 84.1%. Yet had the placement been based solely on the combined cutoff scores for the two tests, with no interpretation (Table 3), the concordance rate would have decreased to 66.5%; decreasing still further if based on only one of those individual tests (Table 3).

Placement into Writing Courses. One hundred, ninety-one of the participants were placed into an ESAL writing course. As indicated in Tables 2 and 3, using the scores adjusted by faculty approximately doubled the accuracy for placing students in the writing courses.

Placement into Communication Courses. One hundred and sixty-eight of the participants were placed into at least one communication course. Of those, 57 enrolled in two upper level communication courses offered concurrently. As illustrated in Table 2, with faculty interpreting the combined scores of the *ACCUPLACER ESL Listening and Language Use* test, and incorporating an oral interview score, course placement was 81.3% accurate. Although any single score would have lowered the accuracy, the oral interview score alone would have been fairly accurate at 62.2%, as would have been the *Language Use* score at 64% (Table 3). However, if the *Listening* test were the sole basis for placing students, the accuracy of the course placement would have dropped to 46.2% (Table 3).

Discussion

This study revealed the overall accuracy of the ESAL course placement process at TRU to be high (81% to 84%). The data presented reveals that extensive faculty involvement—in terms of interviewing the students, assessing the writing sample, and interpreting the diagnostic test results—is imperative for the effectiveness of placement. Without it the accuracy of student placement would decrease significantly.

Of the various skill components, placement into reading courses, based on the combined *ACCUPLACER ESL Reading Skills* and *Sentence Meaning* tests, emerges as the most accurate. According to this data, *Sentence Meaning*, which focuses on vocabulary, is the better indicator of the two. However, clearly a combination of the two scores, balancing the one ability against the other, provides the most effective means of placing students into the reading courses. For the communication courses, again faculty involvement is crucial. However, the independent score for syntax (*Language Use*) seems fairly effective (64%) for placement purposes as does the faculty facilitated oral assessment (62.2%). The *Listening* test score, on the other hand, does not (46.2%). This suggests that if an institution were unable to use the whole battery of *ACCUPLACER ESL* tests, one test that could be excluded would be

Table 2

Placement Based on Interpreted Test Scores vs. Actual Performance in ESAL Reading, Writing, and Communication Courses

Course	Test(s)	Observed	n	Match*	Placement Does not Match**	% Correct	Overall % Correct
Reading	ACCUPLACER ESL Reading Skills and Sentence Meaning	Success	141	136	5	96.5%	84.1%
		Nonsuccess	23	21	2	8.7%	
	Combined Test Scores						
Writing	ACCUPLACER WritePlacer ESL	Success	171	152	19	88.9%	81.2%
		Nonsuccess	20	17	3	15.0%	
Communications	ACCUPLACER ESL Listening and Language Use, and Oral Interview scores	Success	188	178	10	94.7%	81.35%
		Nonsuccess	37	32	5	13.5%	

*Match. Student enrolled in course that was recommended by ESL faculty.

**Does Not Match: Student enrolled in a course higher or lower than recommended by ESL faculty.

Table 3

Placement Based on Test Scores With No Interpretation vs. Actual Performance in ESAL Courses

Placement Course	Test(s)	Observed	n	Match*	Does not Match**	% Correct	Overall % Correct
	ACCUPLACER ESL Reading Skills	Success	141	80	61	56.7%	52.45%
		Nonsuccess	23	17	6	26.1%	
	ACCUPLACER ESL Sentence Meaning	Success	141	84	57	59.6%	58.5%
		Nonsuccess	23	11	12	52.2%	
	ACCUPLACER ESL Reading Skills and Sentence Meaning Combined Test Scores	Success	141	104	37	73.8%	66.5%
		Nonsuccess	23	18	5	21.7%	
Writing	ACCUPLACER WritePlacer ESL	Success	171	74	97	43.3%	44.5%
		Nonsuccess	20	9	11	55.0%	
Communications	ACCUPLACER ESL Listening	Success	188	91	97	48.4%	46.2%
		Nonsuccess	37	24	13	35.1%	
	ACCUPLACER ESL Language Use	Success	188	125	63	66.5%	64.0%
		Nonsuccess	37	18	19	51.4%	
	Oral interview	Success	188	130	58	69.1%	62.2%
		Nonsuccess	37	27	10	27.0%	

the *Listening* test, particularly if the EAP assessment was augmented by an oral interview.

Without faculty corroboration, the *ACCUPLACER WritePlacer ESL* test emerges as least accurate. Out of the 210 electronically scored essays, 117 (56%) had scores that were adjusted by the faculty, in most cases lowered by one level; occasionally by two. The consensus among TRU ESL faculty is that the *Intellimetric* electronic scoring system recognizes discrete features of writing, such as use of semicolons (whether correct or not), and certain more formal transition words (regardless of context). However, the system does seem sophisticated in its capacity to ignore careless spelling, and conversely, to catch senseless rephrasing of the language from the prompt. Despite its limitations, the *WritePlacer ESL* test contributes to the efficiency of the assessment process for a variety of reasons such as enabling the writing component to be integrated in one testing session and providing faculty markers with typed composition samples. Also, students claim to prefer writing in response to the electronic prompt and on keyboards.

In this study the major placement error involved students being placed into courses in which they did not succeed. It is worth noting the type of nonsuccess involved in these cases. For example, in the reading courses, 21 students were predicted to succeed and yet did not. Of these, seven did not complete (DNC), and one withdrew (W). It is possible that external factors may have contributed to the failure of these noncompleters. It is also important to note that this study examined success/nonsuccess by course registrant rather than student count. An example of the implications of this study design is that seven of the noncompleters completely withdrew from the program, receiving DNCs or Ws in all their courses. These seven individual students thus generated 18 placement errors; yet the reason for any of their withdrawals may have been wholly unrelated to academic readiness. Future studies that deliberately focus on this group of noncompleters could yield a clearer picture of such factors.

Although the placement performance of *ACCUPLACER ESL without faculty involvement* appears to be mediocre, it is important to consider that this may have more to do with the complexity of the EAP placement process at TRU than with the testing system itself. For example, based on results from other validity studies assessing the correspondence between automated essay scores and human scores, the correlation ($r=.7$) and the level of agreement for adjacent scores (88%) revealed in this study would be considered quite good and definitely comparable to those reported in such studies (Attali & Burstein, 2005; Burstein, Chodorow, & Leacock, 2004; Elliot, 2003; James, 2006; Landauer, Laham, & Foltz, 2003; Page, 1994; Powers, Burstein, Chodorow, Fowles, & Kukich, 2002; Sireci & Rizavi, 2000; Vantage Learning, 2001). However, for the EAP placement procedure in this study, a one-point difference was significant because it equated to another level in terms of

course placement. This may not be the case for other institutions where EAP programs have fewer levels and/or fewer courses; these institutions may find the ACCUPLACER ESL tests more accurate for their placement purposes.

Moreover, the non-normal distributions of the test scores are more probably due to the dynamics of the EAP population attending TRU at the time of this study rather than any anomalies with the testing tool. Specifically, TRU was experiencing a steady decline in enrollment for the lower levels of EAP, hence the negatively skewed distributions of many of the assessment scores.

Conclusions

The effectiveness of the comprehensive EAP placement process at TRU was determined to depend to a significant degree on extensive involvement by faculty. By facilitating the oral interview, assessing the writing sample, and interpreting the *ACCUPLACER ESL* results, faculty significantly improved the accuracy of student placement. However, given that the total *course* placement error in this study was 18%, refinement of the process and practices used by the ESAL department may be justified. Specifically, the factors contributing to noncompletion merit further investigation, as does the validity of the *ACCUPLACER ESL* tests in general. Moreover, it may be worthwhile to explore the effect of including alternate measurements such as self-assessments as recommended by Rees (1999) or aptitude testing as advocated by Clapham (2000). Based on this study, we highly recommend that multiple measures be employed in *any* EAP placement process and that regardless of the tools employed, faculty be involved in interpreting the results. As Spolsky (1997) stated, "we must make sure that gatekeeping processes remain under responsible and challengeable human control and not be relegated to the automatic and conscienceless power of mechanical self-propelling tests" (p. 247).

The Authors

Cindy James has over 22 years of postsecondary administrative and instructional experience and currently is the Coordinator of the Assessment Centre and a sessional instructor in the Mathematics and Statistics Department at Thompson Rivers University. She also is a doctoral candidate in the Department of Educational Studies at UBC. Her research agenda focuses on issues related to student success and/or psychometrics.

Elizabeth Templeman has taught ESL at Thompson Rivers University, previously chaired the ESL department, and also co-chaired British Columbia's ESL Articulation Steering Committee. Currently she is coordinating an academic support program. She has published book reviews and personal essays in various literary journals and *Notes From the Interior*, a book of creative nonfiction.

References

Alderson, J.C., Krahnke, K.J., & Stansfield, C.W. (Eds.). (1987). *Reviews of English language proficiency tests.* Washington, DC: Teachers of English to Speakers of Other Languages.

Attali, Y., & Burstein, J. (2005). *Automated essay scoring with E-rater v.2.0.* (Research Report No. 04-45). Princeton, NJ: Educational Testing Service.

Ayers, J.B., & Quattlebaum, R.F. (1992). TOEFL performance and success in a master's program in engineering. *Educational and Psychological Measurement*, *52*, 973-975.

Bachman, L.F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing*, *17*(1), 1-42.

Brown, J.D. (1997). Computers in language testing: Present research and some future directions. *Language, Learning and Technology, 1*(1), 44-59. Retrieved September 2006 from: http://llt.msu.edu/volnum1/brown/default.html.

Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated essay evaluation: The Criterion Online Writing Service. *Artificial Intelligence Magazine*, *25*(3), 27-26.

Chalhoub-Deville, M., & Deville, C. (1999). Computer adaptive testing in second language contexts. *Annual Review of Applied Linguistics*, *19*, 273-299.

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

Chapelle, C.A., Jamieson, J., & Hegelheimer, V. (2003). Validation of a Web-based ESL test. *Language Testing*, *20*(4), 409-439.

Clapham, C. (2000). Assessment for academic purposes: Where next? *System*, *28*, 511-521.

College Board (2007). *ACCUPLACER ESL*. New York: Retrieved January 2007 from: http://www.collegeboard.com/student/testing/accuplacer/accuplacer-esl-tests.html

Dunkel, P. (Ed) (1991) *Computer-assisted language learning and testing: Research issues and practice.* New York: Newbury House.

Dunkel, P. (1999). Considerations in developing or using second/foreign language proficiency computer-adaptive tests. *Language, Learning and Technology, 2*(2), 77-93. Retrieved January 2007 from: http://llt.msu.edu/vol2num2/article4/index.html

Elliot, S. (2003). IntelliMetric™: From hear to validity. In M.D. Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 71-86). Mahwah, NJ: Erlbaum.

Fulcher, G. (1999). Assessment in English for academic purposes: Putting content validity in its place. *Applied Linguistics*, *20*(2), 221-236.

James, C. (2006). Validating a computerized scoring system for assessing writing and placing students in composition courses. *Assessing Writing*, *11*(3), 167-178.

Johnson, P. (1988). English proficiency and academic performance of undergraduate international students. *TESOL Quarterly*, *22*, 164-168.

Kahn, A.B., Butler, F.A., Weigle, S.C., & Sato, E.Y. (1994). *Adult ESL placement procedures in California: A summary of survey results. Adult ESL Assessment Project*. Sacramento, CA: California State Department of Education.

Landauer, T.K., Laham, D., & Foltz, P. (2003). Automatic essay assessment. *Assessment in Education*, *10*(3), 295–308.

Light, R.L., Xu, M., & Mossop, J. (1987). English proficiency and academic performance of international students. *TESOL Quarterly*, *21*, 251-261.

Page, E.B. (1994). Computer grading of student prose, using modern concepts and software. *Journal of Experimental Education*, *62*(2), 127–142.

Paltridge, B. (1992). EAP placement testing: An integrated approach. *English for Specific Purposes*, *11*, 243-268.

Person, N.E. (2002). *Assessment of TOEFL scores and ESL classes as criteria for admission to career and technical education and other selected Marshall University graduate programs.* Unpublished master's thesis, Marshall University. (ERIC Document Reproduction Service No. ED 473 756)

Powers, D.E., Burstein, J.C., Chodorow, M.S., Fowles M.E., & Kukich, K. (2002). Comparing the validity of automated and human scoring of essays. *Journal of Educational Computing Research*, *26*(4), 407-425.

Rees, J. (1999). Counting the cost of international assessment: Why universities may need to get a second opinion. *Assessment and Evaluation in Higher Education*, *24*(4), 427-438.

Roemer, A. (2002). A more valid alternative to TOEFL. *College and University*, *77*(4), 13-17.

Sawaki, Y. (2001). Comparability of conventional and computerized tests of reading in a second language. *Language, Learning and Technology*, *5*(2), 38-59. Retrieved January 2007 from:http://llt.msu.edu/vol5num2/sawaki/default.html

Seaman, A., & Hayward, L. (2000). *Standardized ESL test equating study: Equating the CELSA and the NYSPT with the MELT SPLs*. New York: Institute for Education and Social Policy (ERIC Document Reproduction Service No. ED 445084)

Sireci, S., & Rizavi, S. (2000). *Comparing computerized and human scoring of students' essays.* New York: College Board. (ERIC Document Reproduction Service No. ED463324)

Spolsky, B. (1997). The ethics of gatekeeping tests: What have we learned in a hundred years? *Language Testing*, *14*(3), 242-247.

Stricker, L. (2002). *The performance of native speakers of English and ESL speakers on the computer-based TOEFL and GRE General Test.* (Research Report # 69). Princeton, NJ: Educational Testing Service.

Taylor, C., Jamieson, J., Eignor, D., & Kirsch, I (1998). *The relationship between computer familiarity and performance on computer-based TOEFL test tasks.* (Research Report # 61). Princeton, NJ: Educational Testing Service.

Vantage Learning. (2001). *RB-612—Validity issues in automated essay scoring: A summary of four studies.* Yardley, PA: Author.

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

Appendix A
ACCUPLACER ESL Test Descriptions

Reprinted with the College Board's permission from the ACCUPLACER Web site (College Board, 2007) and from the ACCUPLACER Student Guide (College Entrance Examination Board, 2002).

ACCUPLACER ESL Reading Skills

The *Reading Skills* test assesses the student's comprehension of short passages. There are 20 questions to this test each containing a brief passage of 50 words or less, or a moderate length passage of 50 to 90 words. There is a variety of passage content such as the arts, human relationships, physical science, history/social sciences and practical situations. Half of the Reading Skills test contains straightforward comprehension items (paraphrasing, locating information, vocabulary on a phrase level, and pronoun reference). The other half assesses inference skills (main idea, fact vs. opinion, cause/effect logic, identifying irrelevant information, author's point of view, and applying the author's logic to another situation).

ACCUPLACER ESL Language Use

The *Language Use* test measures grammar and usage. It consists of 20 questions in two formats: completing a sentence by filling in a blank with the word or phrase from the choices given, and choosing a sentence that best combines two discrete sentences that are given. The skills covered are subject-verb agreement, verb tenses, forms of ir-

CINDY JAMES and ELIZABETH TEMPLEMAN

regular verbs, appropriate verb forms in structures, noun-noun agreement, noun forms, pronouns, modifiers, comparatives, prepositions, connectives, parallelism, and sentence fragments/run-ons.

ACCUPLACER ESL Sentence Meaning

The *Sentence Meaning* test assesses the understanding of word meanings in one or two-sentence contexts and consists of 20 questions. The sentences are drawn from the content areas of natural science, history/social studies, arts/humanities, psychology/human relations, and practical situations. The areas tested are: particle verbs (or phrasal verbs), basic and important idioms, adverb/adjective phrases, basic nouns, verbs and adjectives (particularly terms of beauty, age, greatness, size), adverbs of frequency, sequence of adverbs (before, after, during), prepositions of direction and place (from, to, at, for, etc), comparatives, connectives, and commands.

ACCUPLACER ESL Listening

The *Listening Test* measures the ability to listen to and understand one or more people speaking in English. The conversations take place in academic environments such as lecture halls, study sessions, a computer lab, the library, the gymnasium, and in everyday environments such as at home, shopping, at a restaurant, at a dentist's office, listening to the radio, reading the newspaper, and performing tasks at work.

For this test, the student listens to the conversation or lecture and the question that follows while looking at pictures of the speakers. Below this are four answer choices, either written on the screen or with a drawing next to each choice. The student chooses the best answer to the question that was asked, and clicks on the button next to his choice. The student can click on the appropriate PLAY button if the student wants to listen again to the conversation, the question, or any of the choices. Each of these can be heard only two more times after the initial reading. There are a total of 20 questions in this test.

ACCUPLACER WritePlacer ESL

The WritePlacer ESL provides a direct measure of the writing skills of students who are not native speakers of English. In this test, a specific prompt is provided and students are asked to provide a writing sample that responds to the prompt. This test can be either untimed or timed. The writing sample is scored on the basis of how effectively it communicates a whole message to the readers for the stated purpose. The score is based on the student's ability to express, organize, and support their opinions and ideas, not the position they take on the essay topic. The following five characteristics of writing are considered: focus, organization, development and support, sentence structure and mechanical conventions.

Appendix B
ACCUPLACER WritePlacer ESL Scoring Rubric
Reprinted with the College Board's permission from the ACCUPLACER Coordinator's Guide (College Entrance Examination Board, 2007).

Score   Description

0   No holistic score—This essay has not been given a holistic score or the five dimensional scores because it is either a blank page, incoherent, insufficient (too short to assess), in a foreign language, or off topic.

1   This writing sample attempts to address the topic but is not effective. There appears to be no pattern of organization of the essay. There is little or no evidence of a main idea, and therefore not enough development and support. Although the essay may contain familiar words written correctly, as a whole it contains a limited vocabulary with many errors in usage. The writing sample may be confusing due to errors in sentence structure, word usage, grammar, and mechanics.

2   Although this writing sample responds to the topic, the main idea is incomplete or hard to understand. The organization contains errors in logic and may not present an effective sequence of ideas. There are few or no details provided to support or develop the topic. The vocabulary used in the essay is simple, familiar, and may be inappropriate. The writing sample typically contains simple sentences, possibly with short, familiar phrases. The sentences contain frequent errors in structure, word forms, and syntax. There are generally frequent errors in mechanics.

3   This writing sample may respond to the topic by stating the main idea, and the main idea may be partly developed using lists and generalizations. Some organization is present, although there are occasional errors in logic and little relationship within and between paragraphs. The vocabulary used in the essay is simple and familiar but may be inappropriate. Typically the writing sample will contain some original simple sentences; however, there are significant errors in sentence structure, word usage, syntax, and mechanics.

4   This writing sample is adequate. A main idea is evident, and there is some evidence of development of this idea. The organization of ideas is adequate. The writer's vocabulary is adequate, even though the vocabulary used may be simple and sometimes inappropriate. There is evidence of some control of word usage and grammar as well as sentence structure. The writing sample typically shows reasonably accurate spelling, punctuation, and capitalization, although there may be some errors in mechanics.

5   This is a strong writing sample. The writing sample states a main idea and provides some specific details and examples to support the main idea. Typically the organization is clear and logical and contains relationships within and between paragraphs that are usually clearly signaled. The choice of vocabulary is sometimes inappropriate. The essay generally shows clarity of expression and provides evidence that the writer can write original sentences and can control most kinds of sentence structures and word forms. Spelling and punctuation are usually correct, although there are occasional errors in mechanics.

6   This is a strong writing sample. The topic is effectively addressed, and the main idea is apparent. The writing sample maintains a central focus. The essay is effectively organized with a logical sequence of ideas that is typically sustained within and across paragraphs. The writing sample provides specific details and examples to develop the main idea. The vocabulary used is generally well suited to what is expressed, and any errors in choice of words do not confuse the reader. The essay typically contains a variety of sentence structures, correct word usage, and correct grammatical forms. The writing sample is generally free of mechanical errors.Appendix C
ESAL Oral Interview Scoring Rubric

Score   Description

0-1   Fluency limited to short responses; pronunciation almost unintelligible. Understands only basic questions and remarks, spoken slowly and distinctly.

2   Speaks with considerable effort and hesitates frequently limited to words of high frequency. Errors in grammar, vocabulary and pronunciation are frequent. Responds to questions and remarks directed to him/her, but may require repetition or a slower speed.

3   Speaks with some effort and occasionally hesitates to search for a word. Vocabulary adequate for general conversation but with errors in word choice. Pronunciation understandable but with non-native stress and intonation, grammatical errors, and poor control of some basic structures. Understands conversation on general topics spoken at a normal rate but may require repetition.

4   Speaks easily—rarely hesitates for lack of vocabulary, occasional errors in pronunciation and grammar occur, speech is understandable. Understands most formal and informal conversation if speech is not rapid. Idiomatic usage may prove difficult.

5   Proficiency equivalent to that of an educated native speaker in comprehension, fluency and breadth of vocabulary and idiom. Errors in grammar are rare and pronunciation almost free of non-native accent.