

Assessing the Usefulness of TOEFL Scores for ESL Decision-Making: A Comparison of iBT, ACCUPLACER-LOEP, and Other Measures of Proficiency

Prepared by Russell Frank, PhD, Pasadena City College, with data analysis from PCC Office of Institutional Effectiveness

Research questions

- **Primary questions:** How do *TOEFL and/or other tests of English language proficiency correlate to the ACCUPLACER-LOEP test currently used for placement at PCC? Can English proficiency tests be used reliably and validly to place students in ESL, English, and other PCC GE classes prior to their arrival at PCC?*

Rationale: Among the PACCD Board of Trustees goals for the 2013-14 school year is the goal of increasing international F1 student enrollment from the current 1,100 students to 2,000 F1 students. One of the challenges presented by the potential increase in this segment of the PCC student population is that of planning and predicting the classes and other resources needed by new F1 students before they have arrived on campus. As a way of addressing the need for data to predict F1 students' ability to succeed in particular ESL, English, and other GE classes, Dr. Bell has raised the question of using TOEFL scores and/or other measures of English proficiency to allow students to register early for classes, or at least for the college to predict the number of appropriate classes needed for new students.

- **Secondary question:** *Can the Common European Framework for Reference to Languages (CEFR) be used as a framework for discussing comparisons among proficiency tests, the ACCUPLACER-LOEP, and PCC ESL course activities and outcomes?*

Rationale: One of the possible backwash effects of researching the proficiency and placement tests at PCC comes from the opportunity to review the information about language proficiency provided by students' test scores and its relationship to the language students need to succeed in their ESL and other classes. The TOEFL, IELTS, ACCUPLACER-LOEP, and other tests are diverse in terms of their approaches to testing language and the way these scores are reported, making direct comparisons extremely difficult. Much work has been done, however, relating individual proficiency tests to the external descriptions of language ability suggested by the CEFR. The goal of using the CEFR to look at student performance is to understand how the information about student proficiency levels from different tests relates to the language proficiency required for success in levels of PCC ESL classes and other classes.

Summary of findings & recommendations

Based on an ordinal regression conducted on the proficiency and placement testing scores of 756 international students entering Pasadena City College between Fall 2011 and Fall 2013, Internet-Based TOEFL scores provided relatively strong predictors of ESL placement based on the ACCUPLACER LOEP ($p=.605$, $p<.001$) and IELTS scores provided moderately strong indicators of placement ($p=.565$, $p<.001$). In both proficiency tests, the writing subtest accounted for the greatest effect on placement, and the spoken subtest accounted for a slightly lower effect. If proficiency tests are to be used for placement, they should be combined with other considerations, such as the students' experiences with English-medium education, the number of iterations of placement tests, and so on.

Background to the primary research question

Measures of PCC international students' English language proficiency and abilities are of critical importance to both students and the College for a variety of reasons. An initial assessment of students' language proficiency allows the college to determine if students are sufficiently proficient to benefit from an education at an English-medium college. As a part of the federal government's requirement for PCC to grant F-1 visas to international students, students must provide evidence of English proficiency in one of the following ways:

- Provide scores less than two years old from an approved English proficiency test.
List of approved tests and related scores:

TEST	Min. Score	Subtests included
<i>Paper-based TOEFL (PBT)</i>	450	Reading, writing, listening
<i>Internet-Based TOEFL (iBT)</i>	45	Reading, writing, speaking, listening
<i>IELTS</i>	5	Reading, writing, speaking, listening
<i>iTEP</i>	4	Reading, writing, grammar, listening, speaking
<i>STEP-Eiken</i>	2 nd Grade	Reading, writing, speaking, listening

- Provide proof of completion of the highest level of ESL classes in an ESL school in the U.S.
- Provide proof of attendance at a high school where English is the medium language of instruction. (min GPA 2.0)
- Provide proof of completion of one academic year of education in the US with a GPA of 2.0 or above (only reflects courses taken in the U.S. (Does not include Intensive English Programs/ESL.)

The following chart shows the ways new F1 students demonstrated their English language proficiency as part of the application submitted to the International Student Center:

PCC International students by proficiency type

Proficiency Type	Number	% of Total
Proficiency test (TOEFL, IELTS, etc.)	756	76.4%
Language school highest level	83	8.4%
US High School/College	112	11.3%
Non-US English Medium School	38	3.9%
Total Students	989	100%

Students send a copy of their test score report to the ISC before an I-20 visa can be issued. Test score reports include the overall score, scores on subtests, and the date the test was taken. Test scores can be no more than two years old. From spring 2011 to Fall 2013, students submitted four kinds of tests scores for admission purposes. The most common test was the ETS Test of English as a Foreign Language (TOEFL) test, submitted by 60% of applicants, followed by Cambridge English Testing Group/British Council/IDP test International English Language Testing System (IELTS), ETS Test of English for International Communication (TOEIC), and the Japanese STEP-EIKEN test, sponsored by the Japanese Ministry of Education. The breakdown in test scores is as follows:

English Language Proficiency by Test Scores

Test	Number	% of Total
TOEFL	535	60%
Internet-based	474	48%
Paper-based	61	12%
IELTS	185	19%
TOEIC	27	3%
STEP-EIKEN	9	1%
TESTS Total	756	76.4%

Tests and Post-Admission Placement

Kokhan (2012, etc.) discusses the difference between general proficiency tests and placement tests:

'Placement tests are designed to assess students' level of language ability so that they can be placed in the appropriate course or class. Such tests can be based on aspects of the syllabus taught at the institution concerned, or may be based on unrelated material' (Alderson, Clapham, & Wall, 1995, p. 11). According to Wall, Clapham, and Alderson (1994), placement tests are subdivided into two types. The first type has a proficiency orientation and may be considered an institutional version of a proficiency test such as TOEFL or the Michigan Test. The content of such tests usually has no direct relationship to the content of the language courses into which students are placed. The second type is developed to specifically reflect the nature of language courses. Universities may give such tests to international students at the beginning of their first semester at the university in order to find out if the students need to take any courses targeting the language skills which are necessary to study in a particular academic environment.

Conversely, proficiency tests are developed to show whether students have managed to reach a certain level of language ability (Alderson et al., 1995) but they may fail to fully meet all the specific language needs of a certain program or educational institution. According to Ascher (1990), English proficiency tests do not measure students' active use of English and may inadequately assess how well students will be able to do in an English-speaking academic environment.

Kokhan studied the relationship between the TOEFL and the University of Illinois English Placement Test (EPT), which includes an essay component scored by trained University of Illinois faculty. Krokhan's recommendations, based on her findings, are informative for the current PCC study:

Based on the findings of this research, it is recommended that TOEFL and EPT score users:

(a) Use the TOEFL iBT total or/and writing section scores as an additional reference [emphasis added] when deciding which level a student should be placed into in case of some disagreement between the EPT raters. The results of one-way ANOVA and the post-hoc Scheffe show that the averages of the TOEFL iBT total and writing sections scores are significantly different among all three placement levels.

(b) Avoid making conclusions about the English proficiency of international students based on their TOEFL iBT scores that are about a year old. According to the pattern of correlation between the time of taking TOEFL iBT and the EPT, the most recent TOEFL iBT total and section scores seem to correlate better with the EPT than the scores which are about 50 weeks old.

While PCC does not currently use the ACCUPLACER writing placement scores from WritePlacer, Kokhan's findings related to the time-sensitive nature of TOEFL scores and the extremely limited predictive ability of these tests point to the need for caution in using test scores outside of their intended purposes.

In addition to the questions above, the accuracy of some TOEFL scores is subject to question. The high stakes nature of the test has led to many fraudulent practices. For example, a local TOEFL test administrator this month caught four individual cheating on the TOEFL. In a follow-up interview with one of those caught, a student at the school where the TOEFL was administered, the student admitted that he had paid \$2,000 for answers to the test. The same version of the test is offered on the East and West Coasts of the US on the day of a particular test administration, and the answers are sent to test-takers, who wrote the answers on small pieces of paper kept inside shirt sleeves. In a related case, a PCC student related an anecdote about her friend that studied in Southern California, failed, transferred to a school in Oregon to raise his GPA, flew to Tijuana to take a TOEFL test for which he had obtained answers, and did well enough to

gain readmission to the school in Southern California. While an estimate of the extent of these fraudulent practices would be difficult to obtain, it can be safely assumed that they are widespread among certain student groups.

Test-takers in countries like China also spend significant amounts of money to attend test preparation schools, like the New Oriental School, which is so large that it is listed on the New York Stock Exchange (NYSE: ADR). Students prepare intensively for tests over a one- to three-month period, often significantly raising test scores but not necessarily raising academic English language proficiency. Such practices serve to highlight the limits to which TOEFL scores can be relied on as sole indicators of language ability.

Background to the secondary research question

The Common European Framework for Reference to Languages (CEFR) was developed to allow educators, language program and materials developers, test developers, and so on have a coherent, language-independent framework for reference to language users' abilities. The following chart shows the six levels and descriptors for the CEFR:

level group	A Basic User		B Independent User		C Proficient User	
level	A1	A2	B1	B2	C1	C2
	Breakthrough (Beginner)	Waystage or elementary	Threshold or (intermed)	Vantage (Upper intermd)	Effective Operational Proficiency (Adv)	Mastery (proficiency)
description	<ul style="list-style-type: none"> Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help. 	<ul style="list-style-type: none"> Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need. 	<ul style="list-style-type: none"> Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise while travelling in an area where the language is spoken. Can produce simple connected text on topics that are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans. 	<ul style="list-style-type: none"> Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialization. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options. 	<ul style="list-style-type: none"> Can understand a wide range of demanding, longer texts, and recognize implicit meaning. Can express ideas fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organizational patterns, connectors and cohesive devices. 	<ul style="list-style-type: none"> Can understand with ease virtually everything heard or read. Can summarize information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in the most complex situations.

Because of the pervasive use of the CEFR in Europe and other countries, developers of high stakes have worked with teams of language testing experts to reference these tests to CEFR levels. The TOEFL, IELTS, TOEIC, and STEP-EIKEN cut scores or ranges have been linked to CEFR groups as follows:

Comparison of Various English Proficiency Test Scores and the Common European Framework for Reference to Languages

CEFR	iBT				PBT (approx.)				IELTS				iTEP		Eiken		TOEIC				
	W	S	L	R	W	S	L	R	W	S	L	R	7.5+	6			W	S	L	R	
C2				29				66					7.5+	6							
C1	28	28	26	28	66		60	64					6.5/7.0	5.5-5.9 5.0-5.4	Grade 1		200	200	490	28	
B2	21	23	21	22	59		55	56					5.0/5.5/6.0	4.5-4.9 4.0-4.4	Pre-1		150	160	400	385	
B1	17	16	13	8	55		48	40					3.5/4.0/4.5	3.5-3.9	Grade 2		120	120	275	275	
A2	11	13			46								3.0	3.0-3.4 2.5-2.9	Pre-2		70	90	110	115	
A1		8												0.0-2.4	Grade 3-5		30	50	60	60	
	Scaled Cut Scores				Scaled Cut Scores				Combined Scores							Scaled Cut Scores					
Source	<i>Tanenbaum & Wylie</i>				<i>Converted from iBT cut score based on ETS "Comparison Chart"</i>				<i>Kies</i>				<i>itepexam.com</i>	<i>stepeiken.org</i>	<i>Tannenbaum & Wylie</i>						

This chart in no way suggests a correspondence among the scores of various test scores or levels. It merely provides a summary of work that has been done referencing individual tests to CEFR groups and levels. As noted above differences in test content and approaches make direct comparison extremely complex.

In the same way that test scores can reference CEFR levels and groups, language classes can address language learners at particular levels. The following chart attempts to relate PCC ESL course learning outcomes according to CEFR levels. The cut scores from the ACCUPLACER-LOEP exam have been included, since these are used in placing students into specific ESL classes, but as noted above, this chart does not imply a direct relationship between the CEFR levels, PCC ESL class levels, and the ACCUPLACER test.

CEFR	LOEP/ACCU.*		PCC ESL*
	R	SM	
C2			
C1	117-120	117-120	
B2	105-116	105-116	ESL 33B? ESL 33A?
B1	88-104	88-104	ESL 122? ESL 422?
A2	60-87	60-87	ESL 420?
A1	29-59	29-59	
	Current PCC Cut Scores		
	<i>Frank</i>		<i>Frank</i>

* ACCUPLACER-LOEP cut scores and PCC ESL levels have been given a preliminary impressionistic assignment to CEFR levels.

References

Boston Educational Services. "iTEP Score Proficiency Chart." N.d. Web. 19 Oct. 2013.

ETS TOEFL. "Internet-based Test Score Comparison Chart." *ETS.com*. 20.5. Web. 10 Oct. 2013.

ETS TOEFL. "Test and Score Summary for TOEFL iBT Tests and TOEFL PBT Tests: January 2012-December 2012 Test Data." *ETS.com*. 2013. Web. 10 Oct. 2013.

IELTS.org. "Guidelines for Educational Institutions, Governments, Professional Bodies, and Commercial Organisations." Jun. 2013. Web. 14 Oct. 2013

Kies, Marylin. "A Framework for Analysing and Comparing CEFR-Linked Certification Exams." *Cambridge English Language Assessment, Research Notes* 42 (2010): 41. Web. 14 Oct. 2013.

Kokhan, Kateryna. "Investigating the Possibility of Using TOEFL Scores for University ESL Decision-Making: Placement Trends and Effect of Time Lag." *Language Testing* 29.2 (2012): 291-308.

Linn, R.I.. Linking results of different assessments. *Applied Measurement in Education* 6.1 (1993), 83-102

Mislevy, J. "Linking Educational Assessments. Concepts, Issues, Methods and Prospects." Educational Testing Service, Princeton, NJ

North, B. Linking language assessments: an example in a low stakes context. *System* 28 (2000): 555-577.

STEP-EIKEN.org. "Grades." Eiken Foundation of Japan. 2013 Web. 19 Oct. 2013.

Tannenbaum, Richard J., and Patricia A. Baron. Research Memorandum ETS RM-11-33: Mapping TOEFL ITP Scores onto the Common European Framework of Reference. *ETS.com*. Nov. 2011. Web. 10 Oct. 2013.

Tannenbaum, Richard L., and E. Caroline Wylie. TOEFL iBT Research Report: Linking English-Language Test Scores onto the Common European Framework of Reference: An Application of Standard-Setting Methodology. *ETS.com*. Jun. 2008. Web. 10 Oct. 2013.

Taylor, L. "Issues of Test Compatability." *Research Notes* 15 (2004): 2-5.

Taylor, Lynda. "IELTS, Cambridge ESOL Examinations, and the Common European Framework." *Cambridge English Language Assessment, Research Notes* 18 (2004): 2-3. Web. 14 Oct. 2013.

Appendix B

Pasadena City College IPRO Recommendations

TOEFL (iBT) and IELTS Predicted Accuplacer ESL Placement

L = Listening Test Score

R = Reading Test Score

S = Speaking Test Score

W = Writing Test Score

Spearman correlation between calculated and predicted ESL placement levels:

TOEFL(iBT): +.605, $p < .001$

IELTS: +.565, $p < .001$

Formulas and Placement Ranges

The following formulas are intended to replicate the ESL placement levels provided by Accuplacer for initial placement. To get students' initial placements, enter their four TOEFL (iBT) or IELTS subscales in Listening, Reading, Speaking, and Writing into the appropriate regression formula below. These 4 subscale scores will be multiplied by the amount displayed in parentheses to their immediate left (e.g., the raw Listening subscale score in TOEFL iBT would be multiplied by 0.01139). Sum up all of the resulting weighted subscale scores to get a total score for that student. Locate the total score in the upper and lower score ranges of the table below (middle and right columns). The student's predicted course placement level will be on the left of that row.

$$Y(\text{TOEFL iBT}) = (0.01139)L + (0.01604)R + (0.20171)S + (0.26198)W$$

Course	Lower	Upper
ESL420	0	3.4742
ESL422	3.4743	6.4241
ESL122	6.4242	8.9858
ESL33A	8.9859	12.7670
ESL33B	12.7671	-

$$Y(\text{IELTS}) = (1.11725)L + (0.89699)R + (-0.01140)S + (0.13930)W$$

Course	Lower	Upper
ESL420	0	7.5903
ESL422	7.5904	10.5583
ESL122	10.5584	13.3308
ESL33A	13.3309	18.5891
ESL33B	18.5892	-

NOTE: Although PCC currently accepts the TOEFL(PBT), we strongly recommend against considering TOEFL(PBT) scores in any placement model. According to the Educational Testing Service's web site, the TOEFL (PBT) lacks the new Speaking test module as well as key formatting enhancements, and is currently being phased out¹. Furthermore, our

¹ <http://www.ets.org/toefl/pbt/about>

exploratory attempts to construct a single consolidated placement model from both the TOEFL(PBT) and the standard TOEFL(iBT), using just the Listening, Reading, and Writing subscale scores, produced less than acceptable results. It is our opinion that the two different TOEFL formats represent two radically different tests.