

# The Duolingo English Test and Academic English



Duolingo Research Report DRR-16-01  
June 29, 2016 (6 pages)  
englishtest.duolingo.com/resources

Lynn Ishikawa<sup>\*</sup>, Kelley Hall<sup>†</sup>, and Burr Settles<sup>‡</sup>

## Abstract

We describe a two-year study of the relationship between the Duolingo English Test (DET) and the academic English ability of non-native international students at DePauw University. We found DET scores to be significantly correlated with on-campus faculty assessments of English ability for incoming international students ( $r = 0.62^{***}$  for written ability and  $r = 0.49^{***}$  for oral comprehensibility). We also found that DET scores significantly predict the faculty's decisions to place certain students into academic English support classes. Furthermore, the DET was more strongly associated with faculty assessments than TOEFL® iBT scores. These results were consistent across both cohorts of first-year students studied (2014 and 2015 incoming classes).

## Keywords

Duolingo English Test, ESL, TOEFL, academic English, language testing

*This report refers to an older version of the Duolingo English Test (DET). More recent research and information on the current test can be found [here](#).*

## 1 Introduction

As is true at many other institutions, DePauw University's student population includes multilingual international students who submit English language assessments as part of their admission materials and take other assessments once they arrive on campus. The university is interested in gathering a complete and accurate picture of students' overall English proficiency both for appropriate admission decisions as well as student placement and support.

At DePauw, academic support related to English proficiency includes formal writing instruction, opportunities to work individually with English for Academic Purposes (EAP) faculty, a language lab, various workshops devoted to academic and language-related skills, and a two-week intensive Summer Institute. Students matriculating into the college are encouraged to attend the Summer Institute, and all take a written assessment during orientation to determine placement into academic English support classes (discussed below). While no one type of assistance — including semester-long courses — automatically produces college-level language proficiency, it is important to determine students' needs in order to offer targeted placement and appropriate instruction. In such a context, the need for an accurate measure of English proficiency is clear. DePauw's motivation to study the relationship between its own assessments of English proficiency and the Duolingo English Test (DET), as well as other traditional language testing, resulted from this need for accurate proficiency measurement.

## 2 Data and Method

All incoming international students at DePauw currently undergo faculty-administered written and oral assessments once they

arrive on campus each fall. These assessments are used to determine whether students should be required or recommended to take additional academic English support classes. This practice of conducting faculty assessments for each international student provides a unique opportunity to study the *criterion validity* of the DET and other English language tests. Since most international students submitted TOEFL® iBT scores with their applications, we include these in our analyses for comparison.

### 2.1 Written assessment

Written assessments were administered to all entering first-year international students during International Student Orientation in August (32 students in 2014; 53 students in 2015). Students had 75 minutes to complete the assessment, which required them to read a short *New York Times* editorial and respond in essay form to a written prompt. They were not allowed to use any outside assistance, including dictionaries or mobile phones, and were asked to hand-write their essays in blue books which were provided by the test proctor. Students had been introduced to academic support staff in an earlier orientation session and had been told that the assessment would determine their placement for fall semester into one of two support classes: ENG 110 or ENG 115 (Academic English Seminar I or II, respectively), or directly into their First-Year Seminar without additional support.

The assessment was scored the following day by a group of English department faculty. Two sample assessments had been chosen for norming purposes by the Director of English for Academic Purposes. The raters were introduced to the six-point scoring rubric (Appendix A) as well as the general characteristics of the support classes offered to first-year students and guidelines for placement into these courses. After this, they read and scored the two norming essays and discussed the rationale behind their scores in an effort to establish consensus

<sup>\*</sup> Director of EAP, DePauw University, Greencastle IN, USA

<sup>†</sup> Associate Dean of Academic Life, DePauw University, Greencastle IN, USA

<sup>‡</sup> Staff Research Scientist & Engineer, Duolingo, Pittsburgh PA, USA

within the group. Consistency for scores of 2, 3, and 4 were of particular importance, since these scores are key in determining placement into support classes (see Appendix B for specific details).

Each faculty member was then given approximately six essays for scoring and instructed to write their initials beside their score and fold the page so that the next rater would be unable to see it. After scoring, the essays were distributed to another faculty member to be read a second time, again with the score hidden. After the second reading, scores were checked by the Director of EAP. If the scores were in agreement, the placement was considered complete for that student and the essay was set aside. In cases where there was disagreement, the essay was redistributed for adjudication readings. For the 2014 cohort, 13 essays (out of 30, plus the two used for norming) received a third reading, and of those two received a fourth reading. For the 2015 cohort, 14 received a third reading, and one received a fourth reading.

## 2.2 Oral assessment

Subsequent oral assessments were administered to first-year international students in September and October. Students were given a link via email to an online schedule and required to sign up for a time slot. The email also explained that they would be asked to take the Duolingo English Test during the same visit (more details on the test in Section 2.3). Assessments were held in the Office of Academic Life on the DePauw campus, administered by the Director of EAP as well as an Academic Coach (in 2014) who had been trained on the assessment practices, and an EAP specialist (in 2015).

Under IRB approval, arriving students were debriefed on the present research study and offered the option to participate;  $n = 77$  agreed (31 from the 2014 cohort; 46 from the 2015 cohort). They were then taken to a testing room in order to complete the web-based DET on a desktop computer, after which they were given an oral assessment interview. (Note that interviewers did not know the results of the Duolingo test until after the interview.) The oral assessment was divided into three sections: familiar topics, longer description, and discussion/analysis. The length of time spent on each section was left to the discretion of the test administrator. For example, in cases where English was a first language, more time may have been spent on questions from the third section (discussion and analysis) than on other sections.

Using a four-point scoring rubric (Appendix C), students were assessed in three aspects of oral communication: comprehensibility, fluency, and pronunciation. The three sections (familiar topics, longer description, and discussion/analysis) were each rated separately, and then averaged to produce a final score for each of the three assessment scores. If a student received a score lower than 3 (out of 4) in any area, he or she received an email recommending regular visits to the Speaking and Listening Center and/or the Director of EAP, as well as attendance at on-campus workshops on class discussion.

Two observations about the oral assessment are worth noting. First, unlike the written assessment, the oral assessment was not intended to be used for placement into academic English support classes, but rather to help identify students who might benefit from additional assistance in aspects of oral communication. Second, an ancillary purpose of the oral assessment is to provide an opportunity for the Director of EAP and other staff to check

in with these students during their first semester of college. In other words, in addition to assessing oral proficiency, the interview provides a valuable opportunity to get to know students and thereby provide more meaningful support throughout their college career.

## 2.3 Duolingo English Test (DET)

In 2014, Duolingo — creator of the online language-learning software — launched the Duolingo Test Center as a novel online language certification platform (<https://englishtest.duolingo.com>). Test Center is designed to provide secure and accessible language assessments, on-demand, to anywhere in the world via Web and Android or iOS mobile devices. Tests are reviewed by remote human proctors using the device's built-in camera and microphone.

The Duolingo English Test (DET) is the first language test offered through the Test Center platform. It is a computer-adaptive test of general English language ability offered for around US\$50. One benefit of computer-adaptive tests is that they can provide uniformly precise scores for the full range of test-takers by administering more difficult items to high-ability students and more easy items to low-ability students. Such tests can typically be shortened by 50% or more (compared to a fixed, non-adaptive version) and still maintain this higher level of precision [1]. The DET usually lasts from 10–25 minutes, including interactive test item formats that take advantage of modern speech technologies, as well as the camera, microphone, and other sensors now available on even the least expensive computers, phones, and tablets around the world. Certificate scores are reported on a 100-point scale<sup>1</sup>. Tests are proctored and results returned in less than 48 hours (median 18.5 hours). DePauw University contacted Duolingo in 2014, shortly after the release of Test Center, to discuss ways of integrating the test into their academic process.

## 2.4 TOEFL® iBT

The TOEFL (Test of English as a Foreign Language) is a standardized test to measure the English ability of non-native speakers, generally used by American universities like DePauw for international admissions. The TOEFL iBT (Internet-based test), launched in 2005, is the preferred TOEFL format today. It is comprised of four subtests — Reading, Listening, Speaking, and Writing — which are scored on a 30-point scale and combined for a 120-point total score. The tests are scheduled at physical testing centers up to four weeks in advance, at an average cost of US\$200 [2]. The Speaking and Writing tests<sup>2</sup> involve the student responding in spoken or written form to several prompts, each of which is scored by a combination of human and/or automated raters [3]. The TOEFL iBT lasts 4–5 hours and results are typically available in two weeks.

## 2.5 Data analysis

We computed descriptive statistics for DePauw faculty assessments, Duolingo (DET) scores, and TOEFL scores, as well as the Pearson and Spearman correlation coefficients among them. All computations were done using the R statistical computing platform [4]. To evaluate how well Duolingo and TOEFL scores might predict faculty placement decisions into support classes, we performed receiver operating characteristic (ROC) analysis [5] using the pROC package [6]. All data were anonymized

by DePauw University's Director of Institutional Research before any analysis took place.

### 3 Results

#### 3.1 Descriptive statistics

Table 1 presents the mean ( $\mu$ ), standard deviation ( $\sigma$ ), and range for DePauw's written and oral assessments, DET scores, and TOEFL scores. Faculty assessments and DET scores are available for all  $n = 77$  students in the study. For various reasons, 16 students did not submit TOEFL scores as part of their application, thus  $n = 61$  for all calculations involving TOEFL scores.

**Table 1.** Summary of assessment results.

Assessment	$\mu$	$\sigma$	Range
Written	2.9	0.95	1–5
Oral Comprehensibility	3.2	0.53	2–4
Oral Fluency	3.2	0.60	2–4
Oral Pronunciation	3.2	0.59	2–4
DET	67.7	19.38	21–100
TOEFL Total	91.7	7.63	77–113
TOEFL Writing	23.2	2.65	17–29
TOEFL Speaking	21.7	2.52	15–28

Note:  $n = 77$ , except  $n = 61$  for TOEFL scores

The mean DET score of 67.7 corresponds to a TOEFL Total score of 95 (according a previous concordance study [7]), which is slightly higher than the actual mean TOEFL of 91.7. Standard deviation and range columns also show a much greater spread among DET scores. TOEFL scores in the sample population are well above average<sup>3</sup>.

Data for the 2015 cohort also include gender, age and nationality. The 2015 international population was 57% female, with an average age of  $\mu = 19.4$  years ( $\sigma = 1.1$ ). 54% were from China, 8% from Vietnam, 6% each from Japan and Pakistan, and the remainder from other parts of Asia, Africa, and Latin America. These demographic variables showed no significant effects on any of the assessments in Table 1, except for a mildly significant difference in the faculty written assessment (3.4 among males vs. 2.6 among females,  $p = 0.02$  using a Welch two-sample  $t$ -test).

#### 3.2 Correlation among assessments

Table 2 shows a correlation matrix of all the associations among assessments, in terms of both Pearson  $r$  (linear) and Spearman  $\rho$  (rank) correlation coefficients. We report both types of correlation because linearly-related scores may not necessarily produce similar rankings (and vice versa); considering one but not the other can be misleading [8].

**Faculty Assessments.** Not surprisingly, the faculty oral assessments — which measure comprehensibility, fluency, and pronunciation — are more correlated with each other (0.64–0.76) than any one of them is with the written assessment (0.32–0.42). This is expected, since the oral assessments are measuring similar constructs of English use, while the written assessment measures a different construct.

**DET Scores.** Since the Duolingo English Test was designed to measure overall English proficiency, we expect it to be

associated with all four faculty assessments. Indeed, DET scores do significantly predict all four assessments, using both correlation measures. Even though all coefficients are statistically significant ( $p < 0.001$ ), the magnitude suggests that the Duolingo construct is most strongly related to written skills and oral comprehensibility.

**TOEFL Scores.** TOEFL Total scores are somewhat associated with the written and oral comprehensibility assessments, and appear unrelated to the other two oral assessments. Instead, it seems more reasonable to compare the TOEFL Writing and Speaking subtests to the corresponding written and oral faculty assessments. The TOEFL Writing test does more strongly predict the written assessment while being unrelated to the oral assessments, and vice versa for the TOEFL Speaking test. The data also suggest that the TOEFL Speaking test is mainly associated with pronunciation.

Note that most of the coefficients are in the 0.3–0.6 range, which is lower than the 0.5–0.7 range we often expect to see in language testing research [9]. This is likely due to the *truncated sample* problem: the students who were given the “gold-standard” faculty assessments had already been filtered by admissions decisions, which was made in part due to their TOEFL scores. Thus, the sample population is inherently biased toward higher scores with a restricted range<sup>4</sup>. This effect artificially deflates correlation coefficients. It is not uncommon for test developers to be satisfied with a coefficient as low as 0.3 with such biased samples [9].

It is also worth noting that the correlation between DET and TOEFL Total scores is only 0.41 ( $p = 0.001$ ), which is lower than the 0.67 ( $p < 0.001$ ) previously reported by Ye [7]. This is likely another artifact of the truncated sample, and of the small sample size generally: the present study includes 77 subjects, whereas Ye's study included 214 (and from more diverse backgrounds). Interestingly, though, DET scores are more highly correlated with the faculty assessments than with TOEFL, while TOEFL Total scores are more correlated with DET scores than the faculty assessments.

#### 3.3 Predicting faculty assessments

If we view the faculty assessments as “gold standards” of readiness for college-level work in English, then the correlation of DET or TOEFL scores with each assessment can be interpreted as a *criterion validity coefficient* for these tests. That is, a higher correlation with the faculty assessment implies stronger validity evidence for the use of the test to assess readiness for college-level work in English.

Comparing the DET and TOEFL validity coefficients in Table 2, DET scores have the highest coefficients with respect to each faculty assessment — using both correlation methods — in seven out of eight cases<sup>5</sup>. The one exception is the TOEFL Speaking test, which seems more linearly associated ( $r$ ) with oral pronunciation. In terms of ranking ( $\rho$ ), however, DET is the most predictive for all four faculty assessments. Table 3 shows that these results are consistent across the two incoming first-year cohorts: DET scores are consistently and significantly associated with all four faculty assessments for both years. Note that the  $p$ -values are lower due to reduced sample sizes, but the overall trend is the same.

**Table 2.** Correlation matrix of all assessments ( $n = 77$ ): DePauw University faculty (W, OC, OF, OP), DET score (D), and TOEFL scores (T, TW, TS). Pearson  $r$  and Spearman  $\rho$  are both reported (above and below the diagonal, respectively). For each faculty assessment, the strongest coefficient among Duolingo and TOEFL scores is highlighted in bold.

Assessment	$\rho \searrow r$	W	OC	OF	OP	D	T	TW	TS
Written	W	—	0.42***	0.41***	0.32**	<b>0.62***</b>	0.41**	0.54***	0.08
Oral Comprehensibility	OC	0.37***	—	0.66***	0.70***	<b>0.49***</b>	0.38**	0.09	0.38**
Oral Fluency	OF	0.39***	0.64***	—	0.76***	<b>0.41***</b>	0.15	0.08	0.29*
Oral Pronunciation	OP	0.34**	0.71***	0.76***	—	0.39***	0.11	-0.04	<b>0.42***</b>
DET	D	<b>0.59***</b>	<b>0.47***</b>	<b>0.37***</b>	<b>0.41***</b>	—	0.41**	0.47***	0.24
TOEFL Total	T	0.29*	0.28*	0.05	0.12	0.34**	—	0.64***	0.47***
TOEFL Writing	TW	0.52***	0.02	0.05	0.02	0.45***	0.65***	—	0.27*
TOEFL Speaking	TS	0.03	0.39**	0.27*	0.35**	0.21	0.44***	0.26*	—

Note: \*\*\* $p < 0.001$  \*\* $p < 0.01$  \* $p < 0.05$

**Table 3.** Detailed validity coefficients by first-year cohort, in terms of Spearman  $\rho$  (rank correlation). For each faculty assessment, Duolingo is the strongest predictor, highlighted in bold.

Faculty Assessment	2014 Cohort ( $n = 31$ )				2015 Cohort ( $n = 46$ )			
	DET	TOEFL	Writing	Speaking	DET	TOEFL	Writing	Speaking
Written	<b>0.69***</b>	0.12	0.47*	—	<b>0.58***</b>	0.36*	0.53***	—
Oral Comprehensibility	<b>0.40*</b>	0.27	—	0.37	<b>0.52***</b>	0.26	—	0.40*
Oral Fluency	<b>0.45*</b>	-0.09	—	0.13	<b>0.39**</b>	0.13	—	0.36*
Oral Pronunciation	<b>0.45*</b>	0.13	—	0.33	<b>0.42**</b>	0.10	—	0.36*

Note: \*\*\* $p < 0.001$  \*\* $p < 0.01$  \* $p < 0.05$

### 3.4 Predicting academic English support needs

Based on the written assessment, first-year international students are either placed into a writing course for non-native English speakers, or are exempt from these academic English support classes and placed directly into their First-Year Seminar (see Appendix B). Another way to evaluate *criterion validity* is to consider how well an English test can predict whether or not university faculty decide to place a student into academic English support classes.

To do this, we use receiver operating characteristic (ROC) analysis [5], a common way to evaluate diagnostic tests in psychology, psychometrics, medicine, and other fields. For our data, the ROC curve illustrates how well a test can predict faculty decisions, if one were to rank students by test score and vary the “threshold” below which students are placed into support classes based solely on that score. As the threshold decreases, both the true positive rate (fraction of students correctly placed directly into First-Year Seminar) and the false positive rate (fraction of students who need support but scored higher than the threshold) will increase.

Figure 1 shows the ROC curves for all students (left), as well as for the 2014 and 2015 cohorts separately (middle and right, respectively). The area under the ROC curve (AUC) — also called the Wilcoxon rank-sum test — is also reported in the legend. This represents the probability that a random student who went directly to First-Year Seminar did in fact receive a higher test score than a random student who was placed into support classes. An “ideal” ROC curve would be right angle in the upper-left corner (perfect true positive rate and zero false positive rate, with  $AUC = 1.0$ ).

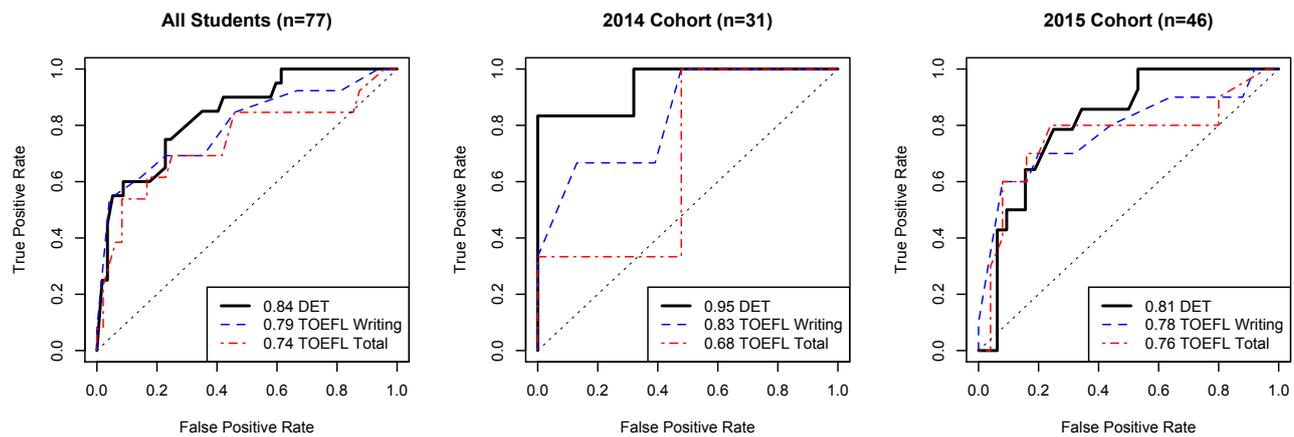
For all three plots, DET scores yield very good ROC curves (AUC > 0.8), and are also consistently better than TOEFL

curves. Closer inspection of the curves reveals that all three scores are similarly good at placing advanced English speakers directly into First-Year Seminar (steep initial lines to the left of each plot). However, DET curves are generally higher in the middle and right side of each plot, indicating that it does a better job assessing the academic English needs of students with medium-to-low proficiency.

## 4 Discussion

Whatever methods are used by colleges and universities to place multilingual students into language and writing classes, there is no doubt that the stakes are high. Indeed, Crusan (2002) refers to such placement as “an act laden with pedagogical, ethical, political, psychometric, and financial implications” (p. 18). In light of this, it seems important that programs obtain the most accurate information available in terms of students’ English proficiency as part of this process. The data presented here provides evidence that the Duolingo English Test — in addition to being more accessible and affordable than other standard English assessments — offers predictive validity in relation to English communication needs in a real university context.

Ultimately, the goal of an assessment is to meet the needs of students. This is certainly true of placement assessments designed to match students to courses that will offer instruction in the skills they need for college-level academic work. Research in academic English writing placement suggests that using multiple assessments — as opposed to a single grammar test, multiple choice test, standardized test, or essay test, for example — may facilitate this process [10]. The findings presented here provide support for the use of the Duolingo English Test as one of the means for placing students into the appropriate course to meet their needs.



**Figure 1.** Receiver operating characteristic (ROC) curves for predicting academic English support needs among all students (left), the 2014 cohort (middle), and the 2015 cohort (right). Area under the curve (AUC) for each test is reported in each graph's legend. DET scores are consistently better than either TOEFL Writing or TOEFL Total scores.

### Limitations

Our findings were consistent across two separate cohorts of incoming international students, which makes a strong case for the validity and replicability of our results. However, the study was conducted among the population of students at a single university. Furthermore, the study design itself limits the ability to draw firm conclusions that compare DET and TOEFL for assessing academic English. While DET scores were consistently more correlated with faculty assessments than were TOEFL scores, DET tests were administered concurrently with those assessments, while TOEFL tests were administered months earlier. This lag time likely reduces TOEFL predictive validity somewhat, as students may have continued (or ceased) studying English in the interim. As indicated in the main text, there is also a truncated sample problem and small sample size generally, as all students were partially admitted on the basis of TOEFL scores. Despite these limitations, our results do strongly indicate that the DET is at least as valid as TOEFL at predicting on-campus academic English ability.

### Notes

- Initially, DET scores were reported on a 10-point scale. In September 2015, Duolingo switched to a 100-point scoring system (using a simple multiplier of 10; the scales are otherwise identical).
- Since the faculty assessments are focused on written and oral skills, we omit the Reading and Listening subtests from our analyses here.
- For comparison, the average 2014 TOEFL scores for all worldwide test-takers were: Total 80.0, Writing 20.3, and Speaking 20.2 [11].
- The above-average TOEFL scores and narrow spread in Table 1 provides additional evidence of truncation.
- Note that TOEFL coefficients are computed using data from 61 subjects, since some students did not submit TOEFL scores. All other correlations use 77 subjects. Results are statistically identical, however, if all analyses are restricted to the 61 students with TOEFL scores.

### References

- [1] D.J. Weiss and G.G. Kingsbury. Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 21:361–375, 1984.
- [2] TOEFL Testing Locations. <http://www.ets.org/bin/getprogram.cgi?test=toefl>, Retrieved February 2016.

- [3] TOEFL iBT® Test Scores. <https://www.ets.org/toefl/ibt/scores/>, Retrieved February 2016.
- [4] R Core Team. R: A language and environment for statistical computing. <http://www.r-project.org/>.
- [5] T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.
- [6] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.C. Sanchez, and M. Müller. pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12:77, 2011.
- [7] F. Ye. Validity, reliability, and concordance of the Duolingo English Test. Technical report, University of Pittsburgh, May 2014.
- [8] J. Hauke and T. Kossowski. Comparison of values of Pearson and Spearman correlation coefficients on the same sets of data. *Quaestiones Geographicae*, 30(2), 2011.
- [9] J.C. Alderson, C. Clapham, and D. Wall. *Language Test Construction and Evaluation*. Cambridge University Press, 1995.
- [10] D. Crusan. An assessment of ESL writing placement assessment. *Assessing Writing*, 8:17–30, 2002.
- [11] ETS. Test and score data summary for TOEFL iBT tests: January 2014–December 2014 test data. <http://bit.ly/1XMafoG>, Retrieved February 2016.

### Appendices

#### A Written assessment rubric

The following describes the six-point scoring guide used by DePauw University (revised August 1995).

1. Critically incompetent; grave deficiencies
2. Definitely incompetent; major deficiencies
3. Marginally incompetent; reaching for but not quite achieving competence; serious deficiencies (e.g., the central idea only vaguely suggested, assertions irregularly explained or illustrated)
4. Marginally competent; but with deficiencies (idea only implied, a gap in development, repetitious sentences pattern)
5. Competent overall; only minor deficiencies (e.g., imbalanced some important assertions lack support)
6. Competent overall; few, if any, minor deficiencies

### Bottom Half Criteria

- Central idea: merely suggested or absent
- Overall pattern of development: inconsistent to non-existent
- Paragraph construction: little discernible pattern; inconsistent, inappropriate, irregular or no details and examples
- Sentences: clouding of meaning
- Mechanics: many or repeated errors of spelling, punctuation, grammar

### Top Half Criteria

- Central idea: stated or clearly implied
- Overall pattern of development: discernible beginning, middle, end
- Paragraph construction: discernible beginning, middle, end: generalizations supported with details or examples
- Sentences: clear; words appropriate
- Mechanics: few errors of spelling, punctuation, grammar

## B Academic English support placement guidelines

The following outlines DePauw University's writing assessment placement guidelines for international students (revised August 2014).

- Each essay gets a minimum of two readers.
- Readers grade each essay on a six-point scale (see rubric below).
- A score of 1 or 2 indicates placement in ENG 110, a score of 3 indicates placement in ENG 115, and a score of 4, 5, or 6 indicates placement directly into the First-Year Seminar (see below).
- Placement will be determined based on the following guidelines:
  - If the essay gets a 1 or 2 from two readers, the student will be placed in ENG 110.
  - If the essay gets a 1 or 2 from one reader and a 3 or higher from another, it will receive a third reading for adjudication.
  - If the essay gets a 3 from two readers the student will be placed in ENG 115.
  - If the essay gets a 3 from one reader and a different score from another, the essay will receive a third reading for adjudication.
  - If the essay gets a 4 or higher from two readers, the student will take their First-Year Seminar with no ENG 110 or ENG 115.

A few suggestions for reading the assessments:

- While it is important to read holistically, it is also important to pay close attention to the student's facility with language. In this way scoring the assessment is different from grading. When grading we try to get past sentence-level errors in order to understand the student's point; when scoring we must try to assess whether sentence-level errors will interfere significantly with the student's ability to communicate in writing.
- Be aware that some errors are more problematic than others. Significant global errors include verb tense and sentence structure, etc., while less significant local errors include articles and prepositions, etc.
- Avoid judging the whole essay based on the first few sentences, especially since these sentences sometimes mimic the language of the prompt. Also, watch to make sure the writer is not using phrases from the article without using quotation marks, as this can give a false impression of fluency.
- Look for evidence of some knowledge of the basics of essay writing: organizing ideas into paragraphs, having an argument, offering evidence/examples, answering the full question, etc. This should count in the writer's favor.

- In general, ENG 110 students lack sufficient control of the language to make their points clearly understood. These are students whose language use will likely interfere with their success in most, if not all, of their classes.
- The following are general characteristics of an ENG 110 text:
  - Difficult to read comfortably; errors require the reader to "work"
  - Contains errors in most sentences
  - Requires significant revision and sentence-level editing
  - Writer would benefit both from additional instruction in American academic writing and in language/grammar
  - Falls into the category of "incompetent" on the 6-pt scale description
- The following are general characteristics of an ENG 115 text:
  - Possible to understand, but with some distraction
  - Contains errors, but chiefly of a local nature (articles, prepositions, etc.)
  - Would benefit from additional help with American academic writing but not necessarily intensive language/grammar work
  - Falls into the category of "marginally incompetent" on the 6-pt scale description

## C Oral assessment rubric

The following describes the DePauw University oral proficiency scale, adapted from Wright State University (revised September 2013).

Fluency (The rate of speech, pauses, flow; ability to talk freely)

1. Speech is so halting and fragmentary or has such a non-native flow that communication is severely impeded.
2. Numerous/frequent non-native pauses and/or a non-native flow, causing the speaker to have some trouble communicating freely.
3. Occasional non-native pauses but with a more nearly native flow so that the pauses do not interfere significantly with communication.
4. Speech is smooth and effortless, similar to that of a native speaker.

Pronunciation (The articulation of sounds)

1. Very Frequent phonemic errors and foreign stress/intonation patterns that cause the speaker to be unintelligible.
2. Frequent phonemic errors and foreign stress/intonation patterns that cause the speaker to be difficult to understand.
3. Occasional phonemic errors and foreign stress/intonation patterns, but speaker is intelligible.
4. Few to no non-native pronunciation/intonation errors; speaker is always intelligible.

Comprehensibility (Can the speaker be understood? Includes grammatical accuracy and vocab use)

1. Significant and persistent grammatical and usage errors in the simplest type of speech.
2. Frequent errors/rephrasing and limited grasp of vocabulary impede comprehensibility.
3. Occasional errors in grammar, word choice, or vocabulary, with some rephrasing, but overall comprehensible.
4. Comprehensible and accurate with impressive control of language with only very minor errors.