OJED
OPEN JOURNALS IN EDUCATION

The Effects of Linguistic and Demographic Features of Chinese International Students on Placement Test Levels in Higher Education: Logistic Regression

Eunjeong Park[a]✉

Abstract: Higher education institutions in the United States provide placement essay tests to ensure international students' readiness for college courses. The high-stakes nature of placement tests makes educators and researchers seek significant components of differentiating levels of placement tests. This study investigated the prediction of two levels (i.e., low vs. intermediate) of 411 placement test essays written by Chinese international students and examined the influence of linguistic and demographic features on placement test levels through logistic regression. The results showed that the type-token ratio, tokens, college type, and graduate status were significant indicators to differentiate students' placement test essays. However, several demographic features were not statistically significant. The results may shed light on improving writing skills of Chinese international students who scored intermediate or low in the placement tests.

Keywords: logistic regression, placement essay test, type–token ratio

Introduction

An influx of international students has been steadily increasing in colleges and universities across the United States. International students are "[second language] students born, raised, and educated in another country who come temporarily to the U.S. on a foreign student visa for a short-term educational or training program" (Ferris, 2009, p. 4; Ferris & Hedgcock, 2013). According to the Open Doors 2017 report, the number of international students studying in the United States grew by 3.4 % in 2016–2017 (Institute of International Education [IIE], 2017). International students constituted approximately 5.3% of students enrolled in the total U.S. higher education student population in

---

✉ Corresponding author: park.1752@osu.edu

the year 2016–2017. Additionally, international students are considered crucial for the economic and social impact on the U.S. International students have contributed more than $35 billion to the economic vigor of American higher education institutions and their communities in 2016-2017. According to the 2017 Open Door report (IIE, 2017), 1,078,822 Chinese international students were enrolled in higher education in the US, making China the top place of origin of international students. International students have also contributed to scientific and technical research and related diverse and dynamic perspectives in U.S. classrooms, benefiting American classmates for global careers and business relationships (IIE, 2017).

Most international students enrolled in U.S. universities and colleges must take placement essay tests to ensure that they are ready for introductory college courses. Placement tests have the purpose of assigning students a specific level of language ability within the curriculum (Brown, 1994; Harmer, 2007; Hughs, 2003). According to the Educational Testing Service (2017), placement tests carry several benefits, such as the incorporation of students' learning to their corresponding proficiency level and the reduction of student and faculty frustration and student retention. Placement test results are also used to assign international students to the appropriate levels of composition classes to meet the academic expectations of institutions. Due to the high-stakes nature of placement testing, concerns about academic performance have generated a considerable interest in the field of educational testing over the years. Furthermore, many linguists, educators, and assessment-related researchers (L. Cheng & Fox, 2017; Read, 2015) have delved into significant components of differentiating levels of placement tests.

The factors of placement test results often influence second language (L2) teaching, such as what to teach in the classroom. Therefore, it is essential to examine what the factors are so that international students receive proper instruction based on their needs. Placement test results can vary due to different factors, such as demographic and linguistic variables (L. Cheng & Fox, 2017; Read, 2000, 2015). Particularly, in this paper, gender, academic status, disciplinary variation, tokens (i.e., the number of words), and type-token ratio (i.e., type-to-token ratio) are examined based on the review of the literature.

Literature Review

Main demographic and linguistic factors were reviewed to better understand the present study.

Demographic Factor 1: Gender Differences in Tests

Many studies (Bible, Simkin, & Kuechler, 2008; Bolger & Kellaghan, 1990; Bridgeman, 1991; Bridgeman & Lewis, 1994) have reported that females tend to do better in essay-type tests than males. Bolger and Kellaghan (1990) investigated the impact of gender on three subjects, namely, English, Irish, and mathematics. The study found that female students performed better on the essay examinations than male students. Bridgeman (1991) and Bridgeman and Lewis's (1994) studies yielded similar results for the Advanced Placement examinations in American and European history, English language, and biology. Female students outperformed male students on the essay examinations in European history and

English language. More studies have also found a positive relation between gender of females and performance on essay examinations (e.g., Becker & Johnston, 1999; Du Plessis & Du Plessis, 2009; Harris & Kerby, 1997; Lumsden & Scott, 1995; Williams, Waldauer, & Duggal, 1992). Oppong's (2013) study also showed that females perform better than males in an essay-type test and explained females' out-performance on essay tests. First, females perhaps do well in novel situations than males. Second, females generally use the English language more often in their communication than males, leading to their ability to perform better in examinations requiring the use of the English language in writing the responses.

While many studies found that females outperform males on essay examinations, other studies found no difference. Lumsden and Scott's (1995) study implied that they did not find a significant gender effect on the economics essay examination of the Chartered Association of Certified Accountants. Holley and Jenkins (1993) did not find a significant gender effect on essay questions. Since gender differences in essay tests are on-going issues, we need to examine if gender differentiates levels of placement test essays, as well as considering other predictors.

Demographic Factor 2: Academic Status as an Indicator of Writing Proficiency

Huang (2010) indicated that international graduate and undergraduate students generally have difficulty in writing essays. Therefore, it is crucial for educators to acknowledge the academic language needs of international graduate and undergraduate students across disciplines. Grover (2013) expressed a concern that there is a need for improving writing skills at the graduate level seemingly due to a lack of preparation of incoming graduate students. A study by Singleton-Jackson, Lumsden, and Newsom (2009) suggested that there is no significant difference in the overall writing skill of undergraduate and incoming graduate students. The findings revealed that the graduate students did not score significantly higher than undergraduate students. However, little research is available on academic status (undergraduate vs. graduate) as an indicator of different levels of writing proficiency. Therefore, this study will contribute to filling a gap in research.

Demographic Factor 3: Disciplinary Variation

Neumann's (2001) conceptual review of quality of university teaching denoted that soft disciplines (e.g., arts) tend to emphasize critical thinking and analysis and synthesis of course contents, whereas, hard disciplines (e.g., science and engineering) tend to emphasize skills in dealing with numerical data with little writing required. Neumann (2001) indicated that soft disciplines receive higher ratings of academic outcomes than hard disciplines. North's (2005a, 2005b) studies revealed disciplinary differences exist related to differing conceptions of the nature of knowledge in a variety of textual features. North's (2005b) study showed that students from an arts background accomplished significantly higher grades of essay writing than those from a science background. The findings of North's (2005b) study suggest that this disciplinary variation presents knowledge by framing the discussion as a matter of interpretation rather than fact.

Furthermore, North (2005a, 2005b) claimed that students' writing is influenced by their disciplinary background. Hyland (2012) also made a contrast between arts and sciences groups by examining distribution and frequency of lexical bundle use. Hyland (2012) stated that argument patterns in the two disciplinary domains (i.e., arts and sciences) are distinctive. The group of business studies and applied linguistics tended to connect aspects of argument in their writing, while the other group of biology and electrical engineering tended to avoid authorial presence by pointing to graphs and findings. As implications from prior literature show (Hyland, 2012; Neumann, 2001; North, 2005a, 2005b), disciplinary variations may represent different levels of language proficiency in placement essay writing.

Linguistic Factor 4: Type–Token Ratio

Vocabulary is an essential component in language learning. Doró (2007) maintained that vocabulary is closely connected to judgment in determining the quality of writing. Singleton (2001) also pointed out that lexical development and vocabulary growth occur in various contexts, particularly in case of a second language. In the field of language research, type–token ratios (TTRs) have been extensively used as an index of lexical diversity (Richards, 1986). Crossley and McNamara (2012) regarded word length, text length, and lexical diversity as "linguistic sophistication" (p. 117) and maintained that text length can be an indicator of differentiating high-proficiency essays from low-proficiency essays.

Read (2000) explained how students with a high TTR use a variety of different words in their writing. The TTR indicates the number of unique words (i.e., type) in comparison to the total number of words (i.e., token) that the writers use with "lexical density." Therefore, lexical density can be measured by the simple TTR. The TTR is a calculation of the number of types divided by the number of tokens in a text with the TTR formula (number of types / number of tokens × 100; Lindqvist, Gudmundson, & Bardel, 2013). Types indicate the unique words in a text, and tokens are subsequent appearances of that word type in the text (Harrington, 2018). The phrase "the big cat in the big hat" has five types (the, big, cat, in, hat) out of seven total tokens. The TTR is a measure of lexical diversity originally developed for measuring first language (L1) vocabulary development. It is an index of lexical diversity and not a measure of absolute size, but it is reasonable to assume that users who produce a wider variety of words—that is, have a higher TTR—will also have larger vocabularies. In practice, however, the measure has been shown to be relatively insensitive to differences in proficiency levels.

The TTR is a helpful measure of lexical variety within a text. Harrington (2018) regarded TTR as an index of lexical diversity, assuming that users who produce a wider variety of words are likely to have a higher TTR. Thomas (2005) explicated that the range falls between a zero (infinite repetition of a single type) and one (the complete non-repetition found in a concordance). The more types there are in comparison to the number of tokens, the greater there is a lexical variety. In other words, a high TTR signifies a large amount of lexical density, while a low TTR implies relatively little lexical density. Read (2000) suggested that students with the lower TTR tend to use a limited number of words

repetitively in their writing. Douglas (2009) also claimed that students with the lower lexically varied essays have less success than the ones with more lexically varied essays.

Linguistic Factor 5: Tokens

Tokens refer to the total number of words. A token is a figure that the word count function of a word-processing program gives, and a type is each repeated item once only (Hunston, 2002). Scholars and linguists have been interested in whether text length can be an effective indicator of writing proficiency. Several studies (Ferris, 1994; Reid, 1986, 1990) have implied that higher-rated essays contain more words. Baba's (2009) study investigated the impact of the lexical proficiency on English language learners' summary writing in English by controlling for the impact of linguistic abilities in English and Japanese (the first language). The participant's English lexical proficiency, English reading comprehension, English proficiency, knowledge of Japanese vocabulary, and writing proficiency in Japanese, and the length of summaries were assessed in the study. Multiple regression analysis of the data showed that reading comprehension and text length were the two strongest predictors of summary writing performance. Crossley and McNamara's (2012) study also examined the importance of lexical variables in writing proficiency. The findings regarding linguistic sophistication raised a theoretical implication that lexical variables (e.g., lexical diversity, word frequency, word meaningfulness, and word familiarity) account for 29% of the variance in the regression analysis. This supports Engber's (1995) claim that lexical knowledge is an important aspect of L2 writing proficiency and indicates the importance of lexical richness and variety in assessing L2 writing skills. Crossley, Kyle, Allen, Guo, and McNamara's (2014) study uncovered that indices of text length, lexical sophistication, and syntactic complexity indicate an ability to quickly and easily produce complex text.

From the literature review, the TTR and tokens can be a useful indicator of lexical variations and learners' language proficiency. Because tokens indicate text length and the TTR represents lexical sophistication (Crossley et al., 2014; Read, 2000), it is necessary to investigate if the TTR and tokens predict the students' levels of language proficiency in the corpus data of placement test essays.

Based on reviews of prior research, this study attempts to examine the influence of gender, academic status, college type, TTR, and tokens on placement test levels (i.e., low vs. intermediate) through logistic regression. The research questions are as the following:

1. Does gender predict the placement test levels (low vs. intermediate)?
2. Do TTR and tokens predict the placement test levels (low vs. intermediate)
3. Do gender, academic status, college type, tokens, and TTR predict the placement test levels (low vs. intermediate)?

Research Method

In this study, we examined participants, data sources (i.e., placement test essays), measures (i.e., TTR, tokens, demographic predictors, and placement test levels as the outcome variable), and effect size for data collection and analysis.

Participants

In 2015-2016, 6,153 international students were enrolled at a midwestern university ("2016 Student Enrollment Report", n. d.). Among the international students, approximately 3,690 (60%) who took placement essay tests agreed to the consent forms. Compositions were only included if there were at least 50 representatives of any given demographic category—that is, gender, country, the college type, and the academic status ("Corpus of Learner English," 2016). Compositions were only included with students' consent and if they did not contain any identifying information about the author. Exclusion criteria involved something personally identifiable (e.g., names and family background), non-consenters' products, and the compositions that did not meet the criteria above. With these thresholds, the 411 compositions in the Corpus of Learner English (CoLE) included all Chinese students (235 males, 176 females), of which 224 were undergraduate and 187 were graduate students from the Colleges of Arts and Sciences (39.4%), Engineering (24.1%), or Business (36.5%). One hundred and forty-one undergraduates received the low level, 83 undergraduates received the intermediate level, 87 graduate students received the low level, and 100 graduate students received the intermediate level. Risks to students were minimized by de-identifying the data in each essay before it was uploaded into the corpus. The corpus is further subdivided into three proficiency levels (low, medium, and high), based on a range of placement test essay scores.

Instrumentation

Placement Test Essays

Placement test essays will be used for examining the most frequently used lexical bundles among the sampled population. The written placement test consists of two parts—a summary and a response to an assigned source. Students read and respond to a source—scientific research articles for graduate students and articles of common sense or about daily living for undergraduate students. Basically, the placement test essays measure how well the students summarize and critically think about the source. Then, their essays are assessed with the three criteria: intellectual property, the content of the source, and language. The English as a second language (ESL) composition coordinator explained the three criteria for evaluating the placement test essays. Intellectual property is measured by whether the students acknowledge a source by citing it in an academically appropriate way; if the content is coherent; if the language is relevant to syntactic structures; and word choice. In addition, several learning skills are reflected in the test: responding to a writing prompt; exhibiting basic critical thinking in response to the text excerpt; organizing ideas in texts; supporting the ideas with relevant readings, personal experiences, and other sources of information; and demonstrating competence in the correct usage of vocabulary, grammar, and mechanics. Based on these criteria, the essays are rated by at least two experienced ESL instructors who are specifically trained on the placement test. They are graded with three rating scores—basic, intermediate, and Q (Qualify).

Measures

TTR

The TTR is a measure of vocabulary variation within a written text or a person's speech. The TTRs of the 411 compositions in the CoLE are calculated and interpreted. The TTR ranged from 33.09 to 67.11 (M = 47.24, SD = 5.98), as seen in Table 1. The TTR is shown to be a helpful measure of lexical variety within a text. It can be used to monitor changes in children and adults with vocabulary difficulties.

Tokens

Tokens (i.e., the number of words) are added to see if the length of an essay matters in the study. Tokens refer to the total number of words. The tokens ranged from 130 to 792 (M = 425.14, SD = 111.83) as shown in Table 1.

Table 1. Descriptive statistics of continuous predictors

Variable	M	SD	Min	Max	Skewness	SE of Skewness	Kurtosis	SE of Kurtosis
Tokens	425.14	111.83	130	792	0.32	0.12	0.50	0.24
Type–token ratio	47.24	5.98	33.09	67.11	0.29	0.12	0.33	0.24

Demographic Predictors

The predictors in this study include gender, academic status, and college type. Among the 411 students, 176 (42.8%) were females and 235 (57.2%) were males. 247 students (60.1%) were in the age range of 18–21, and 164 (39.9%) were over 22. Two hundred and twenty-four (54.5%) were undergraduate students, 187 (45.5%) were graduate students. Among the 411 students who responded, 162 (39.4%) reported being in the arts and sciences, 99 (24.1%) reported being in business, and 150 (36.5%) reported being in engineering.

Placement Test Levels as an Outcome Variable

The placement test level is the outcome variable. Only two levels (i.e., low and intermediate) are included because a focus of this study is on placement. High levels are considered qualified, and students who receive high levels are not assigned to any ESL composition classes. The levels are determined according to the three criteria: intellectual property, content, and language issues. Intermediate levels of placement essays were coded as 1; low levels of placement essays were coded as 0. The residual statistics of the two levels (i.e., low vs. intermediate) of the placement test were checked and indicated normal distribution of residuals in Figure 1.
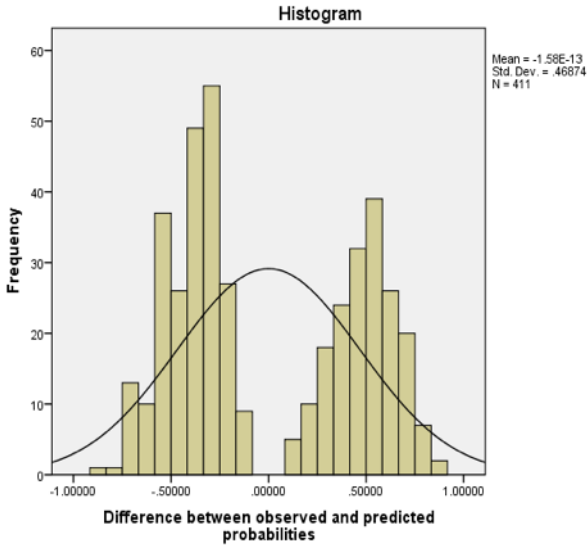
Figure 1. Histograms of the residuals.

Effect Size Statistics

To ensure the results of the data, the effect size was checked in two ways: (a) a post hoc power analysis, and (b) the change of R2 (ΔR2). The logistic regression post hoc power analysis was implemented via G*Power. The one-tailed test yielded the power of .99 under the binomial distribution (α = .05). The change of R2 indicates a weighted least squares effect size measure for the magnitude of the effect. ΔR2 was calculated for each model, such as the difference between the R2 of the null model and the R2 of each model (i.e., Models 1 to 4).

Data Collection

The data were collected from the CoLE. The CoLE is a free and open data set of international students' English placement test essays. The students who took placement compositions at the university testing center consented to having their essays collected into the corpus repository. Demographic information (i.e., gender, college type, and academic status) and placement test levels were included in the corpus of each composition. Compositions were only included with students' consent if they did not contain any identifying information about the author, and if there were at least 50 representatives of any given demographic category.

Data Analysis

To answer the research questions, we employed a logistic regression analysis with the use of SPSS (Version 24; IBM Corp, 2016). Logistic regression is a standard method for finding

the best fitting and the most parsimonious model to describe the relationship between a dichotomous outcome and a set of predictor variables (Hosmer & Lemeshow, 1989). Logistic regression does not require assumptions about the distribution of independent variables. Independent variables do not need to be normally distributed, linearly related to dependent variables, nor have homogeneity of variance across any groups, which is intuitively appealing. To make the intercept meaningful, we dummy-coded all the demographic variables (i.e., gender, academic status, and college) and centered the continuous variables (i.e., the TTR and tokens) around the grand mean.

Binary logistic regression (Hancock & Mueller, 2010; Hosmer & Lemeshow, 1989; O'Connell, 2006) was employed in this study. The first analysis involved gender singly as a binary categorical variable. In the second analysis, the TTR and tokens as continuous variables were included. The TTR was the sole predictor entered first in Model 2-a. Then, tokens were added in Model 2-b. In the third analysis, we included all of the demographic (i.e., gender, academic status, and college type) and linguistic (i.e., the TTR and tokens) variables to examine the statistical significance of them. Finally, the fourth analysis only entered the significant predictors from Model 3. Therefore, Model 4 is a reduced model of the analysis.

Results

This results section includes four different logistic regression models, depending on the interest of predictors.

Model 1: One Dichotomous Predictor (Gender)

Pearson chi-square goodness-of-fit test and likelihood ratio chi-square yielded that the assumption of independence (odds ratio [OR] = 1.0) was not rejected, so there was no significant association between gender and placement test levels as shown in Table 2. Odds of placement test levels were .975 times larger for males than for females. However, the odds were close to the assumption of independence (OR = 0.975, 95% CI [0.658, 1.444]) with no statistical significance. The odds ratio for male versus female students was exp(0.026) = 1.026. An odds ratio is a relative measure of effect by comparing the intervention group of a study to the comparison or control group. If the outcome of the ratio is 1, this implies there is no difference between the two groups of the study. The odds of the intermediate level were about 1.03 times as large for males as females. This indicates that there was no significant association between gender and placement test levels (R^2 = 0.016, df = 1, p = .899).

Model 2: Two Continuous Predictors (TTR & Tokens)

We hypothesized that linguistic features (i.e., TTR and tokens) would provide strong predictors of human judgments of writing proficiency (i.e., low vs. intermediate). In Model 2, we found the effect of the TTR after controlling for tokens as shown in Table 3. The coefficient 0.072 was the log-odds estimate for a one-unit increase in the TTR on the placement test level controlling for the tokens in the model. The coefficient 0.007 was the log-odds estimate for a one-unit increase in the tokens on the placement test level

Table 2. Gender in the equation.

Step 1	B	SE	Wald	df	p	Exp(B)	95% CI for Exp(B) Lower	Upper
Gender	.026	0.201	.016	1	.899	1.026	0.692	1.520
Constant	−.231	0.131	3.088	1	.079	0.794		

Note. Step 1's variable is gender. SE = standard error; Wald = the likelihood ratio test for binary variables; df = degrees of freedom; Exp(B) = the exponentiation of the B coefficient.

controlling for the TTR in the model. The Hosmer and Lemeshow test was not statistically significant (R^2 = 6.138, df = 8, p = .632). Model 2 was improved with the TTR and the tokens by explaining 2.72 % (likelihood ratio R2), 6.5 % (Nagelkerke's R^2), or 8.7 % (Cox & Snell's R^2) of the variance in the placement test levels (i.e., low vs. intermediate). The Wald test suggested that the effect of the TTR and tokens on the log-odds of the placement test level were statistically significant (R^2TTR = 0.072, df = 1, p = .003, 95 % CI [1.024, 1.127]; R^2Tokens = 0.007, df = 1, p < .001, 95% CI [1.004, 1.009]). This indicates that both the TTR and tokens were statistically significant predictors of differentiating placement test levels.

Table 3. TTR & Tokens in the equation

Step 1	B	SE	Wald	df	p	Exp(B)	95% CI for Exp(B) Lower	Upper
TTR	.072	.024	8.667	1	.003	1.074	1.024	1.127
Token	.007	.001	24.419	1	.000	1.007	1.004	1.009
Constant	−.231	.103	5.058	1	.025	0.794		

Note. Step 1's variable are TTR and Token. SE = standard error; Wald = the likelihood ratio test for binary variables; df= degrees of freedom; Exp(B) = the exponentiation of the B coefficient.

Model 3: Additional Predictors (Academic Status, College Type, Gender, Tokens, & TTR)

Model 3 included multiple predictors, such as the academic status, college type, tokens, as well as gender and TTR, although gender and TTR were not statistically significant in Model 1 and Model 2, respectively. The results of Model 3 with six predictors are shown in Table 4. The Hosmer and Lemeshow test in Model 3 was not statistically significant (R^2 = 7.383, df = 8, p = .496), indicating that the model fits the data well. The model explained 8.5% (likelihood ratio R^2), 14.8% (Nagelkerke's R^2), or 11.0 % (Cox & Snell's R^2) of the variance in the placement test levels (i.e., low vs. intermediate). Gender was insignificant with high p values (p > .05). For college type, the College of Engineering was statistically significant (R^2ENG = 7.092, df = 1, p = .008, 95% CI [0.247, 0.8091]), while the College of Business was not significant (R^2BUS = 0.222, df = 1, p = .638, 95 % CI [0.643, 2.055]), given the other variables were held constant in the model. For academic status, the odds

ratio for graduate versus undergraduate students was exp(0.686) = 1.985. The odds of the intermediate level were twice as large for graduate students as undergraduate students. This indicates that there was a statistically significant association between the academic status and placement test levels (R^2 = 3.98, df = 1, p = .046, 95% CI [1.012, 3.893]). The coefficient of tokens, 0.006, was the log-odds estimate for a one-unit increase in the tokens on the placement test level given the other variables were held constant in the model. The Wald test suggested that the effect of tokens on the log-odds of the placement test level was statistically significant (R^2 = 21.45, df = 1, p < .001, 95% CI [1.004, 1.009]). This indicates that there was a statistically significant association between tokens and placement test levels. The coefficient of TTR, 0.064, was the log-odds estimate for a one-unit increase in the TTR on the placement test level given the other variables were held constant in the model. The Wald test suggested that the effect of the TTR on the log-odds of the placement test level was statistically significant (R^2 = 6.516, df = 1, p = .011, 95% CI [1.015, 1.120]). This indicates that there was a statistically significant association between TTR and placement test levels.

Table 4. Demographic and linguistic variables in the equation

							95% CI for Exp(B)	
Step 1	*B*	SE	Wald	*df*	*p*	Exp(B)	Lower	Upper
Gender	.279	.236	1.400	1	.237	1.322	.833	2.098
Col_Bus	.140	.296	.222	1	.638	1.150	.643	2.055
Col_Eng	-.805	.302	7.092	1	.008	.447	.247	.809
AcadStatus	.686	.344	3.980	1	.046	1.985	1.012	3.893
TTR	.064	.025	6.516	1	.011	1.066	1.015	1.120
Token	.006	.001	21.450	1	.000	1.006	1.004	1.009
Constant	-.590	.216	7.454	1	.006	.555		

Note. Step 1's variable are Gender, College of Business, College of Engineering, Academic Status, TTR and Token. SE = standard error; Wald = the likelihood ratio test for binary variables; df= degrees of freedom; Exp(B) = the exponentiation of the B coefficient.

Model 4: Significant Predictors (Academic Status, College Type, Tokens, & TTR)

To develop the most parsimonious and simplistic model, we excluded a non-significant predictor (i.e., gender) and included all the other significant predictors (i.e., college type, academic status, tokens, and TTR) to best fit the data in the logistic model. Model 4 produced similar results to Model 3 with significant predictors as shown in Table 5. The Hosmer and Lemeshow test in Model 4 was not statistically significant (R^2 = 5.333, df = 8, p = .722), indicating that the model fits the data well. The model explained 8.1% (likelihood ratio R2), 14.1 % (Nagelkerke's R2), or 10.5 % (Cox & Snell's R2) of the variance in the placement test levels (i.e., low vs. intermediate). The College of Engineering was statistically significant (R^2ENG = 6.222, df = 1, p = .013, 95% CI [0.271, 0.855]), but the

College of Business was still not statistically significant (R^2BUS = 0.03, df = 1, p = .863, 95% CI [0.597, 1.849]), given the other variables were held constant in the model. For the academic status, the odds ratio for graduate versus undergraduate students is exp(0.94) = 2.559. The odds of the intermediate level were approximately 2.5 times as large for graduate students as undergraduate students. This indicates that there was a statistically significant association between the academic status and placement test levels (R^2 = 13.523, df = 1, p < .001, 95% CI [1.551, 4.223]). The coefficient of tokens, .006, was the log-odds estimate for a one-unit increase in the tokens on the placement test level given the other variables are held constant in the model. The Wald test suggested that the effect of tokens on the log-odds of the placement test level was statistically significant (R^2 = 20.524, df = 1, p < .001, 95% CI [1.004, 1.009]). This indicates that there was a statistically significant association between tokens and placement test levels. The coefficient of TTR, .063, was the log-odds estimate for a one-unit increase in the TTR on the placement test level given the other variables were held constant in the model. The Wald test suggested that the effect of TTR on the log-odds of the placement test level was statistically significant (R^2 = 6.35, df = 1, p = .012, 95% CI [1.014, 1.119]). This indicates that there was a statistically significant association between TTR and placement test levels. The results of logistic regression in Model 4 suggested the following equation:

Logistic regression equation : *Log of odds* = $\ln(p/(1-p))$ = $-.412 + (-.732 * College) + (.94 * Status) + (.006 * Token) + (.063 * TTR)$

Lastly, we considered the assumption of linearity, residual diagnostics, and collinearity. We conducted the Box-Tidwell test to see the assumption of linearity to the logit for the continuous independent variables (e.g., token and TTR). This was done by computing interactions between the continuous predictors and their natural log. The addition of the two product terms was not statistically significant (p = .397). Therefore, linearity in the logit can be safely assumed. We also checked residual diagnostics with influence statistics, such as leverage values and Cook's distance, to identify extreme or unusual cases. Leverage values were less than .029 (3 × 4 [number of predictors] / 411 [sample size]) and the values of Cook's D were also less than 1. Thus, the cases from residual diagnostics may not influence the model's predictions. For collinearity, we ran a regular ordinary least squares (OLS) regression to request collinearity diagnostics. The results showed that tolerance of all the predictor variables was larger than 0.25 and the variance inflation factor was less than 4. Table 6 presents model summaries with coefficients for various logistic regressions.

Discussion

This study implemented logistic regression analyses to make a prediction of the placement test levels. In particular, we will discuss the findings in relation to literature reviews if the results confirm them.

The first research question was whether gender predicts the placement test levels (low vs. intermediate). The results indicated that gender is not a significant predictor for differentiating the placement test levels. As the prior literature (Lumsden & Scott, 1995; Steward, Lim, & Kim, 2015) shows that a significant gender effect of the placement essay test was not found. Therefore, this study did not confirm that females outperform males on

Table 5. Demographic and linguistic variables in the reduced model

Step 1	B	SE	Wald	df	p	Exp(B)	95% CI for Exp(B) Lower	Upper
Col_Bus	.050	.288	.030	1	.863	1.051	0.597	1.849
Col_Eng	−.732	.294	6.222	1	.013	0.481	0.271	0.855
AcadStatus	.940	.256	13.523	1	.000	2.559	1.551	4.223
TTR	.063	.025	6.350	1	.012	1.065	1.014	1.119
Token	.006	.001	20.524	1	.000	1.006	1.004	1.009
Constant	−.412	.171	5.781	1	.016	0.662		

Note. Step 1's variable are College of Business, College of Engineering, Academic Status, TTR and Token. SE = standard error; Wald = the likelihood ratio test for binary variables; df= degrees of freedom; Exp(B) = the exponentiation of the B coefficient.

essay examinations. This would provide a good indication that the placement test results are not be biased by gender.

The second research question was whether both the TTR and tokens predict the level difference. The TTR would be one of the measurement scales of writing because many linguistics researchers (Douglas, 2009; Harrington, 2018; Read, 2000) maintain that the TTR can be a useful indicator of lexical density, variation, and learners' language proficiency. As we examined if both predictors (TTR and tokens) make a difference, it turned out that the TTR and tokens were significant predictors of the placement test levels (low vs. intermediate). With other predictors in Models 3 and 4, the TTR was the significant predictor. As indicated in the Results section, the assumptions of linearity and collinearity of the TTR and tokens were all met. Therefore, it can be said that the single TTR may not be a good predictor to explain the proportion of variance in distinguishing different placement test levels (i.e., low vs. intermediate). Rather, tokens and the TTR together were significant predictors of different placement test levels. This indicates that the length of the placement test essays matters to writing proficiency. The more students write in the placement test, the higher level they can receive. Based on Read's (2000) claim, a high TTR may show lexical variation, which may lead to a higher placement test level. As shown in Model 2, however, the TTR should not be the only predictor of the assessment. The findings from the logistic regression analyses seem to support an evidence of applying the TTR and tokens to assess students' placement test essays.

The third research question was whether all the demographic and language-related variables predict the placement test levels (low vs. intermediate). In Model 3, gender was not statistically significant. As a result, judgment was not made on placement test levels with generic information of gender. Other variables, however, were statistically significant predictors. For instance, graduate students obtained intermediate levels approximately 2.5 times more than undergraduate students. Furthermore, college type was statistically significant. Compared to the College of Arts and Sciences as a baseline, almost half of the students from the College of Engineering received the intermediate test level. North

(2005a, 2005b) indicated that students from an arts background were found to achieve significantly higher grades for essay writing than those from a science background. R. Cheng and Erben's (2012) study also showed that successful language learning was closely tied to the students' programs of study. For instance, students in the art-related programs tended to adapt to a new environment more quickly than those with science-related majors owing to their relatively higher level of productive language skills (e.g., speaking and writing) and frequent interactions with their English-speaking peers.

Several limitations should be addressed for improving the quality of the analysis of models to fit the data. The first limitation is about variable issues. We did not include the information about the advanced levels because the focus of this study is to examine factors of different levels of international students assigned to ESL composition courses. Since we do not have advanced level data, the results may not be generalizable. The second is about the practical significance. Tokens and TTR are significant predictors with a p value ($p <$ .05). However, it should be noted whether the log-odds estimates of 0.006 for the token variable and 0.064 for the TTR variable for a one-unit increase matter to the real world. It would make much sense if the log-odds estimate for a 10-unit increase is assessed to see a practical significance in the data of placement test essay components. Lastly, the variances explained by the models seem limited despite the significant predictors. It may be necessary to explore other predictors for practical value of investigating different levels of placement essay tests.

Conclusion and Implications

This study touched upon what features (i.e., demographic and linguistic) of international students can predict levels (i.e., low vs intermediate) of placement test essays. Gender did not influence the prediction of the essay test levels. Hence, this study made an argument against prior studies (Becker & Johnston, 1999; Du Plessis & Du Plessis, 2009; Harris & Kerby, 1997; Lumsden & Scott, 1987; Oppong, 2013; Williams et al., 1992) about a significant gender effect. Rather, this study agreed with several studies (Holley & Jenkins, 1993; Lumsden & Scott, 1995; Steward et al., 2015) that found no difference based on gender. In order to further study what makes placement test levels different, there is a need to explore potential demographic predictors, such as ethnicity and education in future research.

In light of disciplinary variation (i.e., arts vs. sciences), students from engineering received intermediate levels half as much as those from arts. As Neumann (2001) and North (2005a, 2005b) claimed for the disciplinary variation (i.e., soft disciplines receive higher outcome of academics than hard disciplines), this result may indicate that Chinese international students who are accepted by a certain department, such as engineering, would need more attention to increasing writing proficiency. At an institutional level, departments or colleges should make an effort to support international students by developing and offering appropriate writing courses or workshops based on their academic needs. L2 writing teachers and educators should consider the challenge that the students in hard disciplines (e.g., science and engineering) encounter in writing classes.

As shown in this study, the TTR and tokens are considered as significant indicators of writing levels. The results of this study also affirmed prior literature that length of writing and lexical density would matter in the enhancement of writing proficiency (Baba, 2009; L. Cheng & Fox, 2017; Crossley & McNamara, 2012; Crossley et al., 2014; Doró, 2007; Ferris, 1994; Read, 2015; Reid, 1986, 1990; Richards, 1986; Singleton, 2001). L2 writing teachers also need to consider how they can enhance Chinese international students' writing proficiency by implementing learning strategies of lexical density. Future research should keep in mind these linguistics features and theorize a close relationship for writing proficiency, such as length of writing with tokens and lexical density and variation with the TTR.

References

Baba, K. (2009). Aspects of lexical proficiency in writing summaries in a foreign language. *Journal of Second Language Writing, 18*, 191–208.

Becker, W. E. & Johnston, C. (1999). The relationship between multiple choice and essay response questions in assessing economics understanding. *The Economic Record, 75*(231), 348–357.

Bible, L., Simkin, M. G., & Kuechler, W. L. (2008). Using multiple-choice tests to evaluate students' understanding of accounting. *Accounting Education, 17*, 55–68.

Bolger, N. & Kellaghan, T. (1990). Method of measurement and gender differences in scholastic achievement. *Journal of Educational Measurement, 27*(2), 165–174.

Bridgeman, B. (1991). Essays and multiple-choice tests as predictors of college freshman GPA. *Research in Higher Education, 32*(3), 319–332.

Bridgeman, B. & Lewis, C. (1994). The relationship of essay and multiple-choice scores with grades in college courses. *Journal of Educational Measurement, 31*(1), 37–50.

Brown, H. D. (1994). *Principles of language learning and teaching*. Prentice Hall, New Jersey.

Cheng, L. & Fox, J. (2017). Who are we assessing? Placement, needs analysis and diagnostics. In Cheng, L. & Fox, J., editors, *Assessment in the language classroom: Applied linguistics for the language classroom*, volume (pp., pages 139–165, London. Palgrave.

Cheng, R. & Erben, A. (2012). Language anxiety: Experiences of Chinese graduate students at U.S. higher institutions. *Journal of Studies in International Education, 16*(5), 477–497.

Corp, I. B. M. (2016). IBM SPSS Statistics. *Version, 24.*

Crossley, S. A., Kyle, K., Allen, L. K., Guo, L., & McNamara, D. S. (2014). Linguistic microfeatures to predict L2 writing proficiency: A case study in automated writing evaluation. *The Journal of Writing Assessment, 7*(1).

Crossley, S. A. & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading, 35*(2), 115–135.

Doró, K. (2007). The use of high- and low-frequency verbs in English native and non-native student writing. In Lengyel, Z. & Navracsics, J., editors, *Second language lexical processes: Applied linguistic and psycholinguistic perspectives*, volume (pp., pages 117–129, Clevedon, England. Multilingual Matters.

Douglas, S. R. (2009). Type token ratios and success on the effective writing test.

Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing, 4*(2), 139–155.

Ferris, D. (1994). Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency. *TESOL Quarterly, 28*, 414–420.

Ferris, D. R. (2009). *Teaching college writing to diverse student populations.* Ann Arbor, MI.

Ferris, D. R. & Hedgcock, J. (2013). *Teaching L2 composition: Purpose, process, and practice.* Routledge, New York, NY.

Grover, J. (2013). Writing for the future. *McNair Scholars Research Journal, 9*(1), 21–43.

Hancock, G. R. & Mueller, R. O. (2010). *The reviewer's guide to quantitative methods in the social sciences.* Routledge, New York.

Harmer, J. (2007). *The practice of English language teaching.* Pearson Longman, New Jersey, NJ.

Harrington, M. (2018). Size as a dimension of L2 vocabulary skill. In Harrington, M., editor, *Lexical facility: Size, recognition speed and consistency as dimensions of second language vocabulary knowledge*, pages 3–24, London. Palgrave Macmillan.

Harris, R. B. & Kerby, W. C. (1997). Statewide performance assessment as a complement to multiple-choice testing in high school economics. *Journal of Economic Education, 28*(2), 122–134.

Holley, J. H. & Jenkins, K. E. (1993). The relationship between student learning style and performance on various test question formats. *Journal of Education for Business, 68*(5), 301–308.

Hosmer, D. W. & Lemeshow, S. (1989). *Applied logistic regression.* Wiley, New York.

Huang, L. (2010). Seeing eye to eye? The academic writing needs of graduate and undergraduate students from students' and instructors'. *perspectives. Language Teaching Research, 14*(4), 517–539.

Hughs, A. (2003). *Testing for language teachers.* Cambridge University Press, Cambridge, UK.

Hunston, S. (2002). *Corpora in applied linguistics.* Cambridge University Press, Cambridge, UK.

Hyland, K. (2012). Bundles in academic discourse. *Annual Review of Applied Linguistics, 32*, 150–169.

Lindqvist, C., Gudmundson, A., & Bardel, C. (2013). *A new approach to measuring lexical sophistication in L2 oral production. Eurosla Monographs Series 2: L2 vocabulary acquisition, knowledge and use.* Eurosala, Amsterdam.

Lumsden, K. G. & Scott, A. (1995). Economics performance on multiple choice and essay examinations: a largescale study of accounting students. *Accounting Education, 4*(2), 153–167.

Neumann, R. (2001). Disciplinary differences and university teaching. *Studies in Higher Education, 26*(2), 135–146.

North, S. (2005a). Different values, different skills?: A comparison of essay writing by students from arts and science backgrounds. *Studies in Higher Education, 30*(5), 517–533.

North, S. (2005b). Disciplinary variation in the use of theme in undergraduate essays. *Applied Linguistics, 26*(3), 431–452.

O'Connell, A. A. (2006). *Logistic regression models for ordinal response variables. Thousand Oaks.* SAGE, CA.

Oppong, C. A. (2013). Gender differences in students' performance in history at senior high schools in Cape Coast. *Journal of Arts and Humanities, 2*(1), 34–39.

Plessis, S. D. & Plessis, S. D. (2009). *A new and direct test of the 'gender bias'.*

Read, J. (2000). *Assessing vocabulary.* Cambridge University Press, Cambridge, UK.

Read, J. (2015). Issues in post-entry language assessment in English-medium universities. *Language Teaching, 48*(2), 217–234.

Reid, J. (1986). Using the writer's workbench in composition teaching and testing. In Stansfield, C., editor, *Technology and language testing*, pages 167–188, Alexandria, VA. TESOL.

Reid, J. (1990). Responding to different topic types: A quantitative analysis from a contrastive rhetoric perspective. In Kroll, B., editor, *Second language writing: Research insights for the classroom*, pages 191–210, Cambridge, UK. Cambridge University Press.

Richards, B. (1986). Type/token ratios: What do they really tell us? *Journal of Child Language, 14*, 201–209.

Service, E. T. (2017). Developing an ESL placement test for community college students.

Singleton, D. (2001). *Exploring the second language mental lexicon.* Cambridge University Press, Cambridge, UK.

Singleton-Jackson, J. A., Lumsden, D. B., & Newsom, R. (2009). Johnny still can't write, even if he goes to college: A study of writing proficiency in higher education graduate students. *Current Issues in Education, 12*, 1–39.

Steward, S., Lim, D., & Kim, J. (2015). Factors influencing college persistence for first-time students. *Journal of Developmental Education, 38*(3), 12–20.

Thomas, D. (2005). *Type-token ratios in one teacher's classroom talk: An investigation of lexical complexity (Unpublished master's thesis).* University of, Birmingham.

Williams, M. L., Waldauer, C., & Duggal, V. G. (1992). Gender differences in economic knowledge: An extension of the analysis. *Journal of Economic Education, 23*(3), 219–231.

Author biography

Eunjeong Park is a PhD candidate in the Department of Teaching and Learning, specializing in Foreign, Second, and Multilingual Language Education at The Ohio State University. Her research interests include second language academic writing, instructed second language instruction, and mixed methods research. Email: park.1752@osu.edu