Determining L2 Proficiency in the Production and Perception of Consonant Clusters

PARK, SHINJAE
Department of English Language & Literature
Inchoen National University, South Korea
tlswo@naver.com

ABSTRACT

In recent years, EFL teachers and linguists have evaluated the results of test materials to assess learners' proficiency in languages. However, these approaches have several limitations, such as inappropriate test locations, lack of cost effectiveness, and insufficient test time. To address this issue, in the present study, the author proposes a novel and alternative method of determining L2 proficiency. Specifically, among 6 types of consonant clusters produced by Korean adult L2 learners, specific consonant clusters (e.g., voiceless stop+liquid) related to L2 proficiency were found, through which a series of equations were derived using discriminant analysis; furthermore, the participants' scores in the onset clusters were applied to these discriminant equations to demonstrate the effectiveness of the method; the hit ratio of these equations for categorising L2 level was also examined. In addition, the results of the perception test revealed that the voiced/voiceless stop+glide clusters are related to learners' L2 proficiency. Depending on these induced equations, EFL teachers or researchers can assign a learner to an appropriate proficiency group. In future research, based on different data of EFL learners, the proposed approach can be meaningfully used to derive other useful discriminant equations.

Keywords: consonant clusters; discriminant analysis; L2 proficiency; production; onset

INTRODUCTION

Despite the abundance of previous studies on the relationship between learners' language proficiency and achievement (Garcia-Vasquez et al., 1997; Stevens, Butler & Castellon-Wellington, 2000; Rahman, Yap & Darmi, 2018; Park, 2020), available research has prioritised searching the variables that affect language proficiency. Accordingly, there is a scarcity of studies that have developed mathematical and systematical methods to assess L2 proficiency using quantitative evaluation scores. On the other hand, there is evidence showing that syllable structure is one of the major challenges for EFL learners. Specifically, it is known that speakers of a language that has fewer consonant clusters (e.g., Korean) have difficulties in perceiving and producing a language which has more consonant clusters (e.g., English) (Broselow & Finer, 1991; Eckman & Iverson, 1993; Kwon, 2006; Yun, 2009; Sung, 2018). Kabak and Idsardi (2007), Sung (2018), and Ahn (2020) demonstrated that syllable contact constraints involving L1 phonotactics and L1 syllable structure constraints rather than consonantal contact restrictions caused epenthetic vowels in L2 speech perception. These differences in syllable structures present an obstacle for Korean learners of English. Thus, to evaluate L2 proficiency, the ability to determine whether there is an error of epenthetic vowels in the presented speech which consists of words with word-initial consonant clusters (hereafter, onset clusters), different syllable structures from the Korean language, and the ability to produce them without errors of vowel insertion will be analysed. In addition, to ensure objectivity, the official English test of TOEIC (Test of English for International Communication) is used as a reference to categorise learners' L2 proficiency. The novelty of the present study is that it develops a method to determine Korean EFL learners' proficiency in English by analysing their performance in the perception and production of L2 onset clusters. The effectiveness of the proposed method is also tested. The present study addresses the following four research questions:

(1) Are there differences in how speakers with different levels of L2 proficiency perceive and produce consonant clusters?
(2) Which consonant clusters are most effective in terms of evaluating learners' L2 proficiency?
(3) Is it possible to derive discriminant functional equations using the consonant clusters found in (2)?
(4) If so, to what extent can these equations predict learners' L2 level?

LITERATURE REVIEW

CONSONANT CLUSTERS

There is a broad consensus among scholars concerned with second language acquisition that consonant clusters are a challenge for most L2 learners (Broselow & Finer, 1991; Eckman & Iverson, 1993; Kwon, 2006; Huensch, 2013). Previous studies in consonant cluster acquisition have sought to analyse and explain modification errors made by L2 learners in the production of consonant clusters (Kabak & Idsardi, 2007; Lee & Cho, 2005; Park, 2020). Specifically, several studies demonstrated that L2 learners tend to insert a vowel to break down nonnative consonant clusters (Yun, 2009; Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Park, 2020). Furthermore, several studies reported that Korean learners more frequently make vowel insertion errors in consonant+liquid clusters than in consonant+glide clusters (Kang, 2012; Park, 2020). Likewise, according to Park (2020) [l]-gemination effect results in that, for Korean learners of English, consonant+/l/ clusters are more difficult to pronounce than consonant+/r/ clusters.

The MSD (Minimal Sonority Distance) model, proposed by Broselow and Finer (1991), defines sets of consonant clusters permissible in a language in terms of sonority distances between adjacent segments. According to this model, there is an inverse relationship between the number of clusters allowable in a given language and the degree of sonority distances between adjacent segments. They adopted the sonority hierarchy (Selkirk, 1982) and assigned a sonority value to each class to determine a sonority distance of allowable clusters: Stops (1); Fricatives (2); Nasals (3); Liquids (4); Glides (5).

Accordingly, in languages requiring greater sonority distances between onset consonants, there are fewer types of onset clusters. In contrast, in languages requiring lesser sonority distances, the number of types of onset clusters is larger. The largest possible distance between the sonority index values is 4. Therefore, considering that stops are the least sonorous and glides are the most sonorous, a language with an MSD setting of 5 has no onset clusters. Furthermore, a language with an MSD setting of 1 should theoretically allow all clusters permitted by settings 4, 3, and 2. Accordingly, based on the MSD model, the relative difficulty of production should increase, in the ascending order of markedness as follows: [voiceless stop+glide] < [voiceless stop+liquid] < [voiced stop+glide] < [voiced stop+liquid] < [fricative+glide] < [fricative+liquid]. For instance, the /pj/ sequence is less marked than /pr/, while /pr/ less marked than /fr/.

However, the MSD model was criticised for its reliance on sonority sequencing in determining the difficulty of consonant clusters (e.g., Eckman & Iverson, 1993). Following MSD, given the existence of the onset sequence [nasal+glide] in Korean [myən] 'cotton' and Japanese [myo] 'strange', the Korean and Japanese languages have the MSD setting of 2. Accordingly, consonant clusters with MSD equal to or higher than 2, such as [fricative+liquid] (MSD of 2) and [stop+liquid] (MSD of 3), are to be expected by the sonority distance parameter. Yet, considering that there are no such consonant clusters in Korean or Japanese, Eckman and Iverson (1993) argued that typological markedness is necessary to explain the IL patterns revealed by Broselow and Finer (1991). The definition of typological markedness is as follows:

> Some segments type A is typologically marked relative to some other segments type B if the occurrence of A in a language implies the occurrence of B, but the occurrence of B does not necessarily imply the occurrence of A. Thus, fricatives are marked relative to stops (and equivalently, stops are unmarked relative to fricatives) (e.g., Hawaiian). By similar, now familiar reasoning, voiced stops are marked relative to voiceless stops, and voiced fricatives are marked relative to voiceless fricatives (Eckman & Iverson 1993: 240-241).

However, typological markedness in itself does not tell us anything about certain sequencing constraints holding within syllables. Hence, Eckman and Iverson (1993) used Clements' (1990) Sequential Markedness Principle (thereafter SMP) to predict consonant cluster markedness:

> For any two segments A and B and any given context X_Y, if A is less marked than B, then XAY is less marked than XBY (Clements, 1990: 36).

Eckman and Iverson (1993) examined four Japanese, four Korean, and three Cantonese speakers and claimed that the following markedness aspects would determine the hierarchy of difficulty of clusters for the learners (Eckman & Iverson, 1993): 1) a voiced stop+liquid/glide is more difficult than a voiceless stop+liquid/glide; 2) a voiced fricative+liquid/glide is more difficult than a voiceless fricative+liquid/glide; 3) a voiceless fricative+liquid/glide is more difficult than a voiceless stop+liquid/glide.

Furthermore, according to Kabak and Idsardi (2007), Korean learners perceive that an illusory vowel is inserted between consonant clusters. They believe that there is an illusory vowel between the consonant sequences and that this is directly linked to the vowel insertion in production. Likewise, in a study on the perception of pseudo word 'ebzo', an illegal structure in the Japanese structure, Dupoux et al. (1999) found that most Japanese learners perceived the 'illusory vowel'. In other words, they perceived 'ebzo' as [ebuzo], i.e. relied on their L1 syllable structure rules where such vowel exists even when there is no speech signal in the stimulus. Similarly, in line with Dupoux et al.'s (1999) claim that Korean learners repair illicit onset clusters to adapt to Korean syllable phonotactics by epenthesising fake vowels, Yun (2009) argued that the degree of perceptual epenthesis varied depending on the number and types of onset clusters. Specifically, Yun (2009) reported that illusory vowels were more frequently perceived in the following three conditions: (1) in CCC clusters as compared to CC clusters; (2) in a sequence of obstruents and approximants as compared to a sequence of /s/ and other consonants; (3) after fricatives as compared to after stops in the onset clusters.

RELATIONSHIP BETWEEN ACHIEVEMENT AND L2 PROFICIENCY

Numerous studies have sought to explore whether there is a relationship between L2 proficiency and language learners' cognitive skills, such as general language skills, intelligence, and aptitude (Pratama, Nurkamto, Rustono, & Marmanto, 2017; Tahir, Albakri, Adnan, & Abd Karim, 2020). Language proficiency is defined as the language ability, or ability in language use (Bachmann, 1990; Yazdeli et al., 2016). Likewise, several researchers have proposed to define language proficiency as a language speaker's competency to function in a real language use situation (e.g., Hossein, 1983). Overall, language proficiency has been seen as a multi-componential construct that consists of several sub-constructs (oller, 1983).

There have been several studies on the relationship between students' language proficiency and performance on language proficiency tests. For instance, Stevens et al. (2000) found a strong correlation between the study participants' performance on standardised achievement tests and those students' English language proficiency. Similarly, based on the results of the comparison of Hispanic middle and high school students' proficiency in English and those students' reading achievement scores, Garcia-Vasquez et al. (1997) concluded the highest correlations between English academic achievement and English proficiency (r=0.84).

Likewise, Maleki and Zangani (2007) reported that there the overall performance of EFL learners in language use has a direct and significant impact on those students' English language proficiency.

Interestingly, several studies explored whether students have objective assessment of their level of proficiency in a foreign language. In one such study, Unaldi (2014) compared Turkish EFL learners' self-assessment scores with the results of a placement test and the grades assigned by the instructors. The results of multiple regression analysis showed a strong correlation among all three assessment measures; yet, compared to self-assessment, teachers' assessment appeared to be a much stronger predictor of the students' actual proficiency levels.

However, other researchers, such as Graham (1987), argued that the assumption about the relationship between language proficiency and academic performance has several important limitations. These limitations include inappropriateness of the measures used to define L2 proficiency, ambiguity of the definition of academic success (e.g., GPA may be based on dissimilar courses or on unequal numbers of courses), as well as the possible influence of other variables. In line with this argumentation, Ulibarri et al. (1981) compared 1st, 3rd, and 5th-grade Hispanic students' performance on three English language tests and found that the participants' achievement in reading and math did not correlate with those students' language test data. Similarly, Stevens et al. (2000) compared the performance of English seven-graders on a language proficiency test and a standardised achievement test, and found little correspondence between the results of these two tests. Siding with Graham (1987) and other researchers who questioned the existence of a strong correlation between language proficiency and academic performance, De Avila (1990) argued that learners' acquisition of native-like proficiency in L2 weakens or dissolves the correlation between students' L2 proficiency and academic achievement.

In order to avoid using inappropriate measures to evaluate students' proficiency in L2, this study deems it necessary to use official language test scores. One such test is TOEIC that was originally developed to measure learners' English listening and reading skills. At present, TOEIC is used in government agencies, language schools, academic institutions, and over 4,000 corporations (The TOEIC User Guide, 1999). It is also widely used for business and academic purposes in Korea and Japan. Specifically, Powers, Kim, Yu, Weng, and VanWinkle (2009) found a strong correlation between Japanese and Korean learners' self-reports and those students' TOEIC scores. Specifically, higher TOEIC scores predicted better students' self-assessments in various types of English language tasks. In addition, evidence is available about a correlation between TOEIC speaking and writing measures, suggesting that both measures are relevant to the assessment of EFL learners' proficiency in English (Powers et al., 2009).

METHOD

In the present study, the participants were 55 Korean university undergraduates in the Incheon metropolitan city. Based on the participants' TOEIC scores, they were divided into the following three groups: high proficiency group (HG), n=19; intermediate proficiency group (MG), n=18; and low proficiency group (LG), n=18 (see Table 1). Across all participants, the average TOEIC score was 744.91; the average score of HG, MG, and LG each was 901, 790, and 535, respectively (p<0.05).

TABLE 1. Participants' information

Proficiency	N	TOEIC Range (A)
high group (HG)	19	860~965 (901.05)
intermediate group (MG)	18	690~835 (790.00)
low group (LG)	18	400~630 (535.00)

Total	55	744.91
ANOVA		$F(2, 52)=173.569, p=.00$
Scheffe		LG<MG<HG

A total of 48 nonce words including both phonotactically possible and impossible syllables in English were used as stimuli. The participants performed two tasks: the perception (AXB discrimination) task and the production task. The 48 nonce words contained syllable structures of the following 6 types: 1) voiceless stop+glide; 2) voiceless stop+liquid; 3) voiced stop+glide; 4) voiced stop+liquid; 5) fricative+glide, and 6) fricative+liquid

PERCEPTION

A total of 48 nonce words were used as stimuli. The stimuli were divided into six categories, four consonant clusters per each category (a total of 2640 tokens)[1]

> [voiceless stop+glide]: pj-, pw-, kj-, kw- (e.g., /pyus/, /kwas/)
> [voiceless stop+liquid]: pl-, pr-, kl-, kr- (e.g., /pris/, /klus/)
> [voiced stop+glide]: bj-, bw-, gj-, gw- (e.g., /byus/, /gwas/)
> [voiced stop+liquid]: bl-, br-, gl-, gr- (e.g., /blis/, /grus/)
> [fricative+glide]: fj-, fw-, θj-, θw- (e.g., /fwas/, /thrus/)
> [fricative+liquid]: fl-, fr-, θl-, θr- (e.g., /flus/, /thlis/)

The clusters included both existent and non-existent clusters in English. The basic structure was CCVC (e.g., [θlis], [brus], [fwas]). The structure where a vowel was inserted between consonants was CVCVC (e.g., [θɨlis], [bɨrus], [fɨwas]). In Experiment 1, an AXB discrimination task was performed. The task involved 55 Korean learners who were asked to listen to triplets of words, with two versions according to epenthesis, e.g. [pɨlis]-[pɨlis]-[plis]. They were instructed to choose either the first or the third sound in each triplet as being identical to the second one, and to mark their answers on a sheet of paper. The participants were seated in a soundproof room to listen to the stimuli. The aim was to clarify the perceptual difficulties that Korean learners experience when pronouncing consonant clusters and to find consonant clusters that have a strong effect on their L2 proficiency.

PRODUCTION

The same number of Korean speakers who participated in the experiment 1 also participated in the experiment 2, which was conducted with the same material a week later (total 2640 tokens). The participants read the sentences with 48 target words in a structure of "I say _____ now", and the sound was stored through the 'Tape a Talk' application. Four linguists, including the present author, judged whether or not a vowel (ɨ) was inserted within a cluster in the speech data collected for this study. It is known that Korean learners tend to insert fake vowels to break up onset clusters to adapt to the Korean syllable phonotactics; therefore, vowel insertion is the most frequently occurring error in Korean EFL learners. The first analysis was performed using the 'Praat' software by the present author and one more linguist. In addition, two native English linguists intuitively evaluated speakers' errors. The inter-rater reliability among these four linguists exceeded 80%.

---

[1] 2640 = 55 participants × 48 nonce words. These tokens are also used in the production task.

RESULTS

PRODUCTION

The development of the evaluation method of learners' L2 proficiency by his/her performance on the production of consonant clusters unfolds in the following five steps in line with the process of seeking answers to the research questions raised in the beginning of this study: (1) depending on TOEIC scores, learners are divided into upper-, intermediate-, and lower-level; (2) to verify the multivariate normal distribution, Box's M test is validated; (3) based on the relationship between achievement in production and L2 level, significant variables are selected from among onset clusters; (4) functional equations are formulated; (5) the hit-ratio is validated. Table 2 shows the dependent and independent variables used in the present study.

TABLE 2. Variables and definitions for discriminant analysis

Variables	Definitions
Dependent variables: Categorised by TOEIC scores	
① HG: High proficiency group	· 19 ppl with higher TOEIC score
② MG: Intermediate proficiency group	· 18 ppl with median TOEIC score
③ LG: Low proficiency group	· 18 ppl with lower TOEIC score
Independent variables: onset clusters	
① vl S+G: voiceless stop+glide	· 48 points in total 48 questions
② vd S+G: voiced stop+glide	· 8 points for each variable
③ vl S+L: voiceless stop+liquid	· C1: voiceless stop, voiced stop, fricative
④ vd S+L: voiced stop+liquid	(p,k/b,d/f,ө)
⑤ fric+G: fricative+glide	· C2: glide, liquid
⑥ fric+L: fricative+liquid	(j,w/l,r)

For the statistical analysis, SPSS 25.0 is used. Error frequencies in consonant clusters across the three learner groups are shown in Table 3.

TABLE 3. Group statistics in production

Types		HG error	HG SD	MG error	MG SD	LG Error	LG SD	Total error	Total SD
vl. stop	G	.53	.772	.11	.323	.22	.428	.29	.567
	L	.74	1.33	1.89	1.41	3.11	1.13	1.89	1.61
vd. stop	G	.37	.496	.78	1.17	.67	.970	.60	.915
	L	1.74	1.41	2.67	1.28	3.22	1.67	2.53	1.56
fric.	G	1.05	.705	.89	1.02	1.11	1.13	1.02	.952
	L	2.68	1.95	2.89	1.49	3.89	1.91	3.15	1.84

The results of the analysis of variance (ANOVA) show that differences in the production of the voiced/voiceless stop+liquid cluster across the three groups are significant, providing the answer to the first research question about whether there are differences in performance between groups (see Table 4). According to the results, the voiceless stop+liquid scores (for which Wilks' Lamda[2] score is the smallest, and F-value is the largest) have the

---

[2] Wilks' lambda, which is equal to the proportion of the total variance in the discriminant scores not explained by intergroup differences, is a measure of how well each function separates cases into groups.

highest discriminant power in terms of predicting L2 level, followed by the voiced stop+liquid ($p<.05$).

TABLE 4. Tests of equality of group means

Types	Wilks' Lamda	F	*df* 1	*df* 2	*p*
vlS+G	.901	2.861	2	52	.066
vlS+L	.626	15.529	2	52	.000
vdS+G	.963	.996	2	52	.376
vdS+L	.841	4.908	2	52	.011
Fricative+G	.990	.257	2	52	.775
Fricative+L	.917	2.353	2	52	.105

Table 5 reports the results of the equality assumption[3] of the covariance matrices in each classified group. The data do not differ significantly from multivariate normality (Box's M test=.846, p=.662), so further analysis can be performed.

TABLE 5. Box's M test of equality of covariance matrices

Box's M		.846
	F(Approx.)	.412
F	*df*1	2
	*df*2	6074.071
	p	.662

In order to seek the answer to the research question 2, Table 6 shows the results of the stepwise regression analysis which demonstrates that the groups differed only in the performance of voiceless stop+liquid clusters. This indicates that only one of the two statistically significant variables reported in the analysis of the equality test in Figure 4 can be used as a variable to develop equations in the discriminant analysis in the subsequent sections. This is in line with the result that the value of Wilks' lambda is the lowest in the voiceless stop+liquid in Table 4. Said differently, the value of Wilks' lambda refers to the proportion of the total variance in the discriminant scores not captured by intergroup differences, so lower Wilks' lambda values show a greater discriminatory ability of the function. In the present study's results, the lambda value of 0.626 has a significant p-value (p=0.000) in the voiceless stop+liquid clusters, suggesting the statistical difference among group achievement.

TABLE 6. Valuables entered/removed

Entered	Stat.	*df* 1	*df* 2	*df* 3	Exact F Stat.	*df* 1	*df* 2	*p*
vlS+L	.626	1	2	52.000	15.529	2	52.000	.00

Table 7 shows the canonical correlation coefficient reflecting the degree to which the discriminant score of the voiceless stop+liquid predicts the categorised L2 level. Canonical correlation is used to evaluate the relationship among variables. In this study, it is examined the relationship between the production of the voiceless stop+liquid cluster and L2 proficiency using the canonical correlation. The canonical correlation coefficient amounts .611, suggesting the discriminant power of 37.33% (($.611)^2$=0.3733). Said differently, learners' pronunciation

---

[3] Two fundamental assumptions in Discriminant Analysis are as follows: (1) predictor variables should have a multivariate normal distribution, and (2) within-group variance-covariance matrices should be equal across groups.

scores of accuracy in voiceless stop+liquid cluster have a 37.33% explanatory power on learners' proficiency levels.

TABLE 7. Summary of canonical discriminant

Type	Eigen value	Variance%	Cumulative%	Canonical correlation
vlS+L	.597a	100.0	100.0	.611

Three linear discrimination equations for the three proficiency groups based on the results of the Fisher discriminant functions are shown in Table 8, which gives the answer to the research question 3. Based on these induced three separate equations, EFL teachers or researchers can assign a learner to an appropriate proficiency group. Therefore, applying each learner's performance score for the voiceless stop+liquid variable to the three discriminant equations makes it possible to categorise that learner to the group with the highest score.

TABLE 8. Fisher's linear discriminant functions and equations

Categories	LG	MG	HG
vlS+liquid	1.854	1.126	.439
Constant	-3.983	-2.162	-1.260
Discriminant Equations for LG, MG, HG	DL=1.854 × (voiceless S+liquid)-3.983 DM=1.126 × (voiceless S+liquid)-2.162 DH=0.439 × (voiceless S+liquid)-1.260		

For instance, the error frequency of the top 50% students (median value) in each group is applied to the three derived equations to estimate how accurately those equations allocate learners to L2 proficiency groups. The number of errors in the voiceless stop+liquid clusters of those students is 3 in the low proficiency group, 2 in the intermediate group, and 0 in the upper group (see Figure 1). Figure 1 shows the number of errors for the corresponding onset consonant cluster (i.e., voiceless stop+liquid) of the top 50% participants in each group.
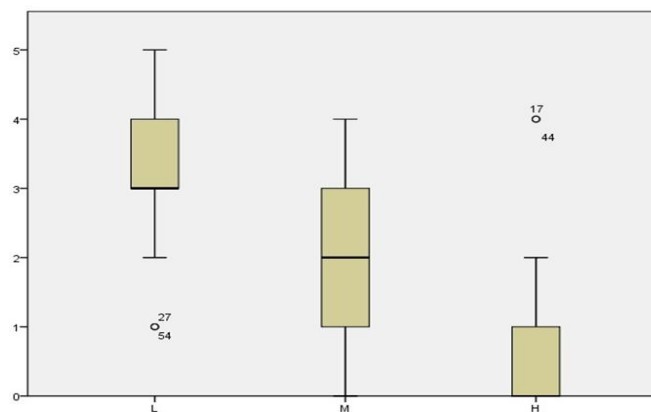

FIGURE 1. Box plots of the results of voiceless stop+liquid

Therefore, entering '3' as the error frequency of the median student in the lower proficiency group yields the following results:

DL=1.854 × (3)-3.983=1.579
DM=1.126 × (3)-2.162=1.216
DH=0.439 × (3)-1.260=0.057

Based on the results of the three equations, it can be determined that the learner belongs to the low-proficiency group (1.579) because the score derived is the highest. This procedure may also apply to median students from two other proficiency groups[4].

The predictability of the above discriminant equations for presenting the answer to the last research question is reported in Table 9. Using the discriminant equations with the voiceless stop+liquid scores in production yields the hit-ratio[5] of categorization, 56.4%. Since the predictive hit-ratio of low proficiency and high proficiency groups amounts to 77.8% and 78.9%, respectively, the accuracy will be dramatically increased up to 78.4% if only two groups are included. Therefore, using the scores in voiceless stop+liquid clusters to categorise learners into three groups yields the accuracy of 56.4%; at the same time, grouping the learners into two groups increased the accuracy to 78.4%.

TABLE 9. Classification results

| Group | Predicted Group Membership | | | Total |
	LG	MG	HG	
Original Count LG (%)	14 (77.8)	2 (11.1)	2 (11.1)	18
MG (%)	8 (44.4)	2 (11.1)	8 (44.4)	18
HG (%)	2 (10.5)	2 (10.5)	15 (78.9)	19

PERCEPTION

Similarly to the analysis of production, in the analysis of perception, this study tries to develop discriminant equations upon establishing which consonant clusters are related to learners' L2 level. The developed equations help to determine which group a learner should be assigned to in the absence of official English scores.

Table 10 shows the results of one-way ANOVA performed for each independent variable. The difference is reported in the results of voiceless stop+glide and voiced stop+glide among the three groups ($p<.05$). Specifically, the groups differ the most (i.e. Wilks' lambda is the smallest and the F value is the highest) at voiceless stop+glide (Wilks' lambda=.724, F=9.908, $p<.05$), followed by voiced stop+glide (Wilks' lambda=.846, F=4.731, $p<.05$).

TABLE 10. Tests of equality of group means

Types	Wilks's lamda	F	*df* 1	*df* 2	*p*
vlS+G	.724	9.908	2	52	.000
vlS+L	.919	2.303	2	52	.110
vdS+G	.846	4.731	2	52	.013
vdS+L	.997	.083	2	52	.921
Fricative+G	.987	.350	2	52	.706
Fricative+L	.938	1.731	2	52	.187

Table 11 reports the equality of the covariance matrices in the classified groups. The results of the multivariate normality test show that the variance of three groups differs significantly ($p<.05$), suggesting that the results violate the following two fundamental assumptions: (1) predictor variables should have a multivariate normal distribution, and (2) within-group variance-covariance matrices should be equal across groups. Accordingly, further

---

4 If we apply the scores of the 50th ranked learner, '2' in MG, the results are as follows: DL=-0.275, DM=0.09, DH=-0.382. Therefore the learner belongs to MG. If we apply the same procedure to the learner's scores in HG, '0', the results are as follows: DL=-3.983, DM=-2.162, DH=-1.260. Therefore the learner belongs to HG.
5 Hit ratio = Accurately Predicted Frequency/Total samples. Hit ratio is a relative concept, so it is hard to find a clear standard for good hit ratio.

analysis cannot be performed. However, in what follows, this study reports the results showing which consonant clusters influences L2 level divided by TOEIC scores.

TABLE 11. Box's M test of equality of covariance matrices

	Box's M	17.743
F	F(Approx.)	5.538
	df1	3
	df2	208080
	p	.001

To find out which consonant clusters are related to L2 proficiency in the perception task, stepwise multiple regression analysis is performed. This approach is selected because it does not need to meet the fundamental assumption of normal distribution (see Table 12).

TABLE 12. Model summary

Entered	Types	R	R^2	Adjusted R^2	Std. Error of the Estimate
step 1	voiceless S+G	.534	.285	.272	142.089
step 2	voiceless S+G & voiced S+G	.586	.343	.318	137.461

When voiceless stop+glide is entered, R^2 amounts to 0.285; furthermore, when voiced stop+glide is also entered in the model R^2 increases to 0.343. This suggests that an increase of the predictability power for L2 proficiency is reported as 0.58 (voiceless S+G: F=21.129, *p*=.000; voiceless S+G: F=13.602, *p*=.000). Accordingly, learners' proficiency can be determined with the accuracy of 34.3% when considering only the scores of voiceless stop+glide and voiced stop+glide in perception. Consequently, discriminant functional equations could not be produced, as the results of Box's M test did not meet the fundamental assumptions; however, the results of the additional regression analysis show that learners' L2 proficiency levels can be predicted by considering just the scores of voiceless stop+glide and voiced stop+glide with the accuracy of 34.3%.

CONCLUSION

In the field of language studies, despite the richness of previous studies on the correlation between official English scores and L2 achievement, prior attempts to develop mathematical and systematical methods to evaluate L2 proficiency level with quantitative evaluation scores have been scarce. To fill this gap in the literature, in the present study explored whether the achievement of perception and production of particular consonant clusters can determine the proficiency of learners in L2. To this end, this study developed a discriminant analysis method to evaluate learners' L2 proficiency with the scores of the statistically derived onset clusters, verified its effectiveness with the actual scores of the consonant clusters by the median-level learner in each group, and analysed the predictive power of the induced discriminant equations. The results show the scores of the voiceless stop+liquid clusters contribute the most to determining the difference between three groups in production. Based on these findings, and using Fisher's linear discriminant function, three discriminant equations were produced to categorise learners' proficiency in L2. Entering the voiceless stop+liquid scores in the derived equations made it possible to estimate learners' proficiency levels with the accuracy of 56.4%. In addition, considering the predictive hit-ratio of the low proficiency group (77.8%) and high proficiency group (78.9%) only, the accuracy dramatically increased to 78.4%.

The results show that the participants' performance when pronouncing the consonant clusters can be applied to the equations to quickly categorise the learners according to their L2 levels. In addition, when learner do not have official language test scores, researchers or educators can categorise learners based on their L2 proficiency with the help of the L2 discriminant functional equations alternatively. This can help educators reduce efforts, such as creating test materials by themselves or arranging test locations and time, and minimise other incidental expenses. Accordingly, the results of the present study provide meaningful insights for EFL teachers and researchers.

On the other hand, in the perception task, since the statistical results did not meet two fundamental assumptions (i.e., predictor variables should have a multivariate normality and within-group variance-covariance matrices should be equal across groups), adequate discriminant functional equations were not established. Yet, the voiceless/voiced stop+glide scores had a predictive power of 34.3% for determining learners L2 proficiency by TOEIC scores. This was performed using a complementary stepwise regression analysis.

In further research, it would be meaningful to analyse those grammatical, lexical, or semantic features of L2 that learners perceive as difficult to acquire. Said differently, future research could extend the present investigation to other linguistic areas, including syntax, semantics, corpus studies, and phonetics (e.g., scores by type of onset cluster or coda cluster). The methods proposed and applied in the present study can help researchers to derive discriminant equations to effectively evaluate learners' proficiency in L2.

ACKNOWLEDGEMENTS

REFERENCES

Ahn, M. (2020). Phonotactics and speech syllabification by Korean learners of English. Studies in Phonetics, Phonology and Morphology, 26(1), 89-101.

Bachman, L. F. (1990). *Fundamental consideration in language teaching*. Oxford: Oxford University Press.

Broselow, E. & Finer, D. (1991). Parameter setting in second language phonology and syntax. *Interlanguage studies bulletin (Utrecht). 7*(1), 35-59. https://doi.org/10.1177/026765839100700102

Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M. Beckman (Eds.), Papers in laboratory phonology I (pp. 283–333). New York, NJ: Cambridge University Press. https://doi.org/10.1017/CBO9780511627736.017

De Avila, E. (1990). *Assessment of language minority students: Political, technical, practical, and moral imperatives*. Washington, DC: U.S. Department of Education.

Dupoux, E., Pallier, C., Kakehi, K., & Mehler, J. (2001). New evidence for prelexical phonological processing in word recognition. *Language and Cognitive Processes. 16*, 491–505. https://doi.org/10.1080/01690960143000191

Eckman, R. F. & Iverson, K. G. (1993). Sonority and markedness among onset clusters in the interlanguage of ESL learners. *Second Language Research. 9*(3), 234-252. https://doi.org/10.1177/026765839300900302

Garcia-Vazquez, E., Vazquez, L. A., Lopez, I. C., & Ward, W. (1997). Language proficiency and academic success: relationship between proficiency in two languages and achievement among Mexican American Students. *Bilingual Research Journal. 21*(4), 395-408. https://doi.org/10.1080/15235882.1997.10162712

Graham, J. G. (1987). English language proficiency and the prediction of academic success. *TESOL Quarterly. 21*(3), 505-521. https://doi.org/10.2307/3586500

Hossein, F. (1983). New directions for ESL proficiency testing. In Oller J.W. (Ed.). Issues in language testing research. (pp. 253-268). U.S.A: Newbury House.

Huensch, A. (2013). The perception and production of palatal codas by Korean L2 learners of English. PhD thesis, University of Illinois, Urbana, USA.

Kabak, B. & Idsardi, W. (2007). Perceptual distortions in the adaptation of English consonant clusters: syllable structure or consonant contact constraints? *Language and Speech, 50*, 23–52. https://doi.org/10.1177/00238309070500010201

Kang, S. Y. (2012). The patterns of vowel insertion in Korean speakers' production of English C+/l/ and C+/r/ clusters. *Phonetics and Speech Sciences. 4(*4), 3–17. https://doi.org/10.13064/KSSS.2012.4.4.003.

Kwon, B. (2006). Korean speakers' production of English consonant clusters: Articulatory and perceptual accounts. PhD thesis, Michigan State University, USA.

Lee, S. & Cho, M.-H. (2005). Perception and production of consonant clusters by Korean learners of English. *English Language and Literature. 50*(5), 1101–1132.

Maleki, A. & Zangani, E. (2007). A Survey on the Relationship between English Language Proficiency and the Academic Achievement of Iranian EFL Students. *Asian EFL Journal. 9*(1), 86-96.

Oller, J. W. (1983). *Issues in language testing research*. Rowley, Mass: Newbury House.

Park, S. (2020). Production pattern of English C+liquid clusters by L2 proficiency of Korean speakers. *English Language and Literature. 25*(3), 299-316. https://doi.org/10.46449/MJELL.2020.08.25.3.299

Powers, D. E., Kim, H., Yu, F., Weng, V. Z., & VanWinkle, W. (2009). The TOEIC® Speaking and Writing Tests: Relations to Test-Taker Perceptions of Proficiency in English. ETS, Princeton, New Jersey. http://dx.doi.org/10.1002/j.2333-8504.2009.tb02175.x

Pratama, H., Nurkamto, J., Rustono, R., & Marmanto, S. (2017). Second Language Learners' Comprehension of Conversational Implicatures in English. 3L: Language, Linguistics, Literature®, 23(3), 50-66. http://doi.org/10.17576/3L-2017-2303-04

Rahman, A., Yap, N. T., & Darmi, R. (2018). The association between vocabulary size and language dominance of bilingual Malay-English undergraduates. 3L: Language, Linguistics, Literature®, 24(4), 85-101. http://doi.org/10.17576/3L-2018-2404-07

Selkirk, E. (1982). The syllable. The structure of phonological representations, 2, 337-383.

Stevens, R. A., Butler, F. A., & Castellon-Wellington, M. (2000). Academic language and content assessment: Measuring the progress of English language learners. Los Angeles: University of California.

Sung, E. (2018). The effects of consonant contact constraints and syllable structure on speech perception in Korean assimilation contexts. Studies in Phonetics, Phonology and Morphology, 24(2), 147-172. http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE07524819

Tahir, M. H. M., Albakri, I. S. M. A., Adnan, A. H. M., & Abd Karim, R. (2020). The Effects of Explicit Vocabulary Instructions on Secondary ESL Students' Vocabulary Learning. 3L: Language, Linguistics, Literature®, 26(2), 158-172. http://doi.org/10.17576/3L-2020-2602-12

Unaldi, I. (2014). Self and teacher assessment as predictors of proficiency levels of Turkish EFL learners. *Assessment & Evaluation in Higher Education. 41*(1). 67-80: https://doi.org/10.1080/02602938.2014.980223

Ulibarri, D. M., Spencer, M. L., and Rivas, G. A. (1981). Language proficiency and academic achievement: A study of language proficiency tests and their relationship to school ratings as predictors of academic achievement. *NABE Journal. 5*(3), 47-79. https://doi.org/10.1080/08855072.1981.10668410

Yazdeli, R. J., Mellati, M., & Mehdizadeh. M. (2016). The relationship between lexical fluency, temporal fluency, and language proficient in Iranian EFL context. *3L: The Southeast Asian Journal of English Language Studies. 22*(1), 111-125. http://doi.org/10.17576/3L-2016-2201-09

Yun, G. (2009). Illusory vowel perception in English onset clusters by Korean learners of English. *Studies in Modern Grammar. 57*, 15–41.