QBS
Vol. 36, No. 2, pp. 119-122 (2017)
https://doi.org/10.22283/qbs.2017.36.2.119
pISSN 2288-1344 eISSN 2508-7185 **Original Article**

Quantitative Bio-Science
Official Journal of the Institute of Natural Science at Keimyung University
Available Online at https://qbs.kmu.ac.kr:442/

A Study on Korean Language Learner Variables based on the Statistical Method

Jincheol Park[1], Kyung-Mo Min[2], Seon-Jung Kim[2,*]

[1]Department of Statistics, Keimyung University, Daegu 42601, Korea
[2]Department of Korean Studies, Keimyung University, Daegu 42601, Korea
(Received October 15, 2017; Revised November 15, 2017; Accepted November 18, 2017)

ABSTRACT

Various factors that exert an influence on foreign Korean-language learners' achievements have long been studied. Recently, to obtain a better understanding of differences in learning achievements, studies have begun comparing learning achievements between learners in the same educational environment. In this study, we examine whether there exists a difference in academic achievements between genders and between *Hanja* and non-*Hanja* culture learners. We also investigate whether those who have received Korean language education from basic to intermediate level with identical textbooks and identical curricula have higher academic achievements than those who have not received such education. To this end, using the linear mixed model, we conducted a statistical analysis of Korean-language fulfillment score data collected from a Korean-language institute administrated by K-university. We found that the high-scoring group tended to make slow, constant progress with regard to advancing to Korean-language level, whereas the low-scoring group tended to withdraw after a significant degree of achievement. Additionally, female learners performed much better than male learners in the high-scoring group, but in the low-scoring group, much poorer performance was observed. Finally, we found that it is a better strategy to start learning Korean from the basic level to the intermediate level at the same university.

Key words : Basic level, *Hanja* culture, Intermediate level, Korean language, Learning achievement, Linear mixed model

Introduction

Factors that influence achievements in Korean language learning include learners' native language, gender, age, time of residence in Korea, personality, proficiency, foreign language aptitude, and learning strategies. In particular, the influence of learner's native language on Korean learning has been discussed by many researchers. In their studies, learners were not only grouped by their native languages–which were English, Chinese, and Japanese–but the characteristics of Korean learning were also examined from the perspective of proficiency, linguistic areas such as grammar and pronunciation, and functions such as listening and reading. Recently, researchers have been conducting more refined studies to understand differences in learning achievements between learners in the same educational environment.

The purpose of this study is summarized as follows.

(1) Do learner variables act differently in groups with high academic achievements or low achievements?

In this study, we examined whether there exists a difference in academic achievements between genders and between *Hanja* culture learners and non-*Hanja* culture learners. Several studies have been conducted in which academic achievements in Korean language education showed that female

* Correspondence should be addressed to Dr. Seon-Jeong Kim, Department of Korean Studies, Keimyung University, Daegu 42601, Korea. Tel: +82-53-580-6941, Fax: +82-53-580-5115, E-mail: kimsj@kmu.ac.kr

learners performed better than male learners, and some studies showed that there were no gender-based differences among the learners. The effect of learners' gender on academic achievements has been analyzed among Korean government scholarship students by measuring their achievements after the completion of Level 1 in [1]. The results showed no correlation between gender and academic achievements. Analyzing the 7th, 8th, and 9th TOPIK (Test of Proficiency in Korean), [2] found that female examinees were rated higher than male examinees in all areas, including vocabulary, grammar, writing, listening, and reading. However, in both studies, only the average scores of female and male learners were compared and analyzed. It has been shown in [3] that *Hanja* culture learners had higher achievements in vocabulary, grammar, and reading in comparison to non-*Hanja* culture learners. However, this study also compared and analyzed only the average scores of the two groups' achievements. Therefore, in the current study, the characteristics of Korean language learning are analyzed by classifying the two groups on the basis of the assumption that students with high achievements and students with low achievements are likely to show different characteristics. Through the analysis, we try to explore whether differences in academic achievements between genders and differences between *Hanja* culture learners and non-*Hanja* culture learners show the same or different characteristics in the high-achievement group and low-achievement group.

(2) Is learning Korean in the same language institute from basic to intermediate level better?

In this study, we compare the learning achievements of learners who have learned Korean from basic to intermediate level in the same institution and learners who have learnt Korean from basic to intermediate level in different institutions. This is to determine whether those who have received Korean language education from basic to intermediate level with identical textbooks and identical curricula have higher academic achievements than those who have not received such education. This new variable, specific to Korean learners, is a novel concept that so far has never been studied in the field of Korean language education.

Methods

We aimed to cluster students of beginning or intermediate Korean language level by investigating the patterns of levels of fulfillment as the students advanced in levels and to characterize the identified groups in terms of gender and membership of *Hanja*/non-*Hanja* culture. To this end, we decided to use the data of students who had registered from 2008 to 2017 at the Korean-language institute of K-university in Daegu city. The K-university provides a quarter system and level-based education in a Korean-language institute where 1∼2 level, 3∼4 level, and 5∼6 level correspond respectively to beginning, intermediate, and advanced levels. For a representative measure of students' level of fulfillment, we use the scores of the exam administered to students to determine whether they have succeeded or failed to advance a level. We selected the records of students who had promoted levels without failing. To answer question (2), mentioned in the Introduction, we selected two data sets, G1 and G2.

The G1 is the data set of students who started learning Korean language and advanced at the same institute from levels 1 to 5 without fail. The G2 is the data set of students who had prior knowledge of the Korean language and started off learning at the institute from levels 2 to 5. The number of students who registered more than 4 quarters was 1055, and the number of available score records was 5029. Further narrowing down the data set by selecting records suitable for analysis, we obtained the records of 210 students (G1: 122, G2:88) in which G1 and G2 respectively consisted of scores of levels 1∼4 and 2∼4. We also imputed missing scores using the multiple imputation technique, which has been proven successful in various applications [4].

For each student i, we observed a score vector $(Y_{i1},Y_{i2},Y_{i3},Y_{i4})^T$ for G1 data set and a $(Y_{i2},Y_{i3},Y_{i4})^T$ for G2 data set, where Y_{ij} denoted the standardized 100 percentile score obtained at level j. Then, employing Partitioning Around Medoids (PAM) [5], an unsupervised clustering technique, we identified two subgroups of students for G1 and G2 data sets in which two clusters were selected as the optimal number of clusters for both G1 and G2. Fig. 1 shows the plot of the average silhouette of each cluster size for the G1 data set.

As we can notice from Figs 2 and 3, PAM clusters students' groups into two clusters where the dashed lines and dotted lines respectively correspond to students of high (cluster = 0) and low (cluster = 1) level of fulfillment.

For the data analysis, we introduced a binary variable "cluster" that denoted the cluster picked up by the PAM algorithm. In addition to clustering information, we considered three factors in explaining the scores: gender, *Hanja*/non-*Hanja* cul-

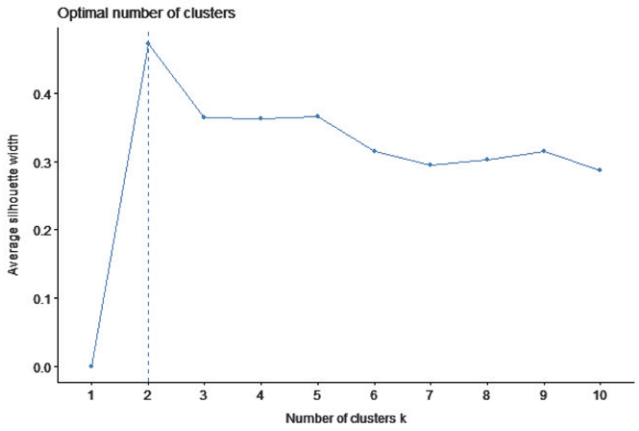

Fig. 1. Plot visualizing the optimal number of clusters using the method of average silhouette for the G1 data set.

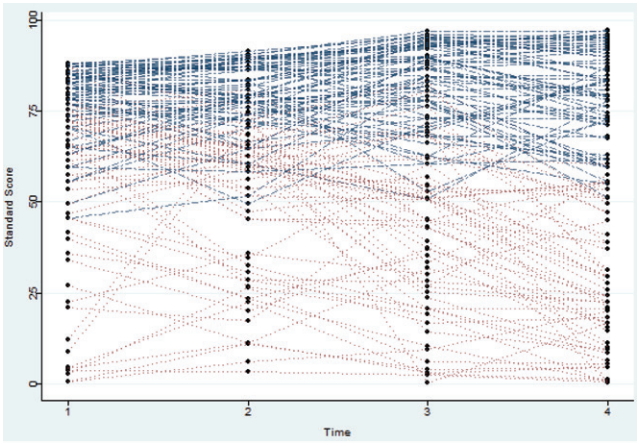

Fig. 2. The longitudinal profile plot of the scores achieved by each student in the data set G1. The x-axis and y-axis denote level and stand score respectively. Membership of each profile line is differentiated by line types (dotted or dashed).

ture, and level of Korean-language. The gender variable g is a binary variable defined to be 1 for females and 0 for males. The binary variable h indicates *Hanja* for zero and non-*Hanja* otherwise. The quantitative l variable denotes student level. Because we have repeatedly measured observations, we employed a linear mixed model formulated by

$$Y_{ij}=\beta_0+A_i+\beta_1 cluster_i+\beta_2 g_i+\beta_3 h_i+\beta_4 l_i+ \beta_5(cluster\times g)_i+\beta_6(cluster\times h)_i+\beta_7(cluster\times l)_i+\varepsilon_{ij}, \quad (1)$$

where $A_i \sim IID\ N(0,\sigma_A^2)$, $\varepsilon_{ij} \sim IID\ N(0,\sigma_\varepsilon^2)$, and A_i are independent from ε_{ij}.

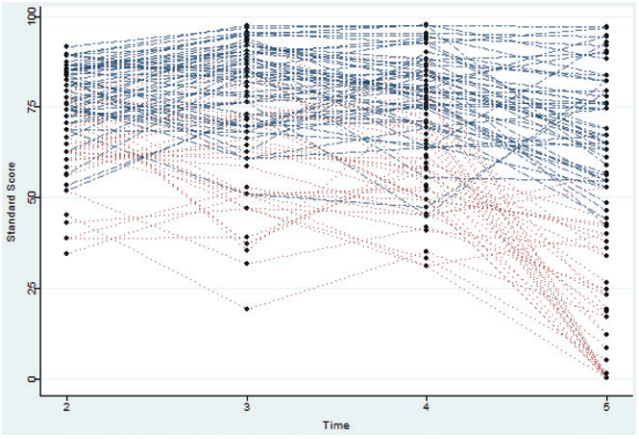

Fig. 3. The longitudinal profile plot of the scores achieved by each student in the data set G2. The x-axis and y-axis denote level and stand score respectively. Membership of each profile line is differentiated by line types (dotted or dashed).

Table 1. The estimates and associated p-values of the parameters of the linear mixed model (1) for G1.

Parameter	Estimates	p-values
β_0 intercept	71.5	<0.0001
β_1 cluster	−8.05	0.081
β_2 gender	5.31	0.095
β_3 *Hanja*/non-*Hanja*	3.52	0.266
β_4 level	0.82	0.207
β_5 cluster × gender	−13.14	**0.005**
β_6 cluster × *Hanja*/non-*Hanja*	16.65	**0.001**
β_7 cluster × level	−10.04	**<0.0001**

Table 2. The estimates and associated p-values of the parameters of the linear mixed model (1) for G2.

Parameter	Estimates	p-values
β_0 intercept	84.6	<0.0001
β_1 cluster	14.75	0.007
β_2 gender	1.83	0.353
β_3 *Hanja*/non-*Hanja*	1.70	0.353
β_4 level	−2.64	0.005
β_7 cluster × level	−11.88	**0.005**

Results and Discussion

We implemented model (1) using lme4 R package [6]. For the G1 data set, the estimates and associated p-values are displayed in Table 1. Now that all the interactions were significant, the discrepancy in the scores of females versus males was significantly different between high- and low-achievement groups. In the case of the high-achievement group (cluster = 0), the females achieved 5.31 points higher scores than

males on the average. Meanwhile, in the low-achievement (cluster = 1) group, females achieved about 7 lower scores than males on average. In the same argument, the discrepancy in the scores between *Hanja* and non-*Hanja* were also significantly different between the high and low-achievement groups. In the case of high achievement (cluster = 0), students of *Hanja* culture achieved a 3.52 higher score than those of non-*Hanja* on average. In the low-achievement (cluster = 1) group, *Hanja* achieved about 20 higher score than non-*Hanja* on average. The interaction between cluster and level also suggested a different performance trend between cluster 0 and 1: the scores of students of cluster 0 tended to increase by 0.82 points, but those of cluster 1 were likely to decrease by 9 points as students advanced one level.

For the G2 data set, we fitted the model (1) and found that only $cluster \times l$ was statistically significant so that model (1) without $cluster \times g$ and $cluster \times h$ interaction effects was finally selected, yielding the estimates and associated p-values in Table 2. Gender and *Hanja* culture effects were not been found to be significant. For level factors, now that the interaction $cluster \times l$ is significant, we can say that cluster 0 and 1 showed different trends in level effect: the scores of the students tended to decrease by 2.64 points as the high-achievement group (cluster = 0) advanced one level, and the scores were even more likely to decrease for the low-achievement group−that is, by 14 points−as students advanced one level.

In summary, in answer to the first research question, we can say that learner variables act differently between high- and low-achievement groups. In particular, the high-scoring group tends to make slow, constant progress in advancing Korean-language levels. Meanwhile, the low-scoring group tends to withdraw after a significant degree of achievement. Another point we would like to make is that females perform much better than males in the high-scoring group, but in the low-scoring group, much poorer performance is observed among females. In an answer to the second research question, after comparing the analysis results of G1 and G2, we can say that it is more effective for students to learn Korean from basic to intermediate level at the same institute.

Acknowledgements

This research was supported by the Keimyung University Research Grant of 2017.

References

1. Choi G. A study on anxiety of Korean language learners in a second language learning environment - on the correlations between the Korean language classroom anxiety and the academic achievement, age, and sex of the postgraduate scholarship students of the Korean government. Journal of Learner-Centered Curriculum and Instruction 2016;16:953-974.
2. Chang E, Yang K. Detecting the differently functioned items by gender on the test of proficiency in Korean. Bilingual Research 2006;32:291-324.
3. Kim S-J, Park J, Min K-M. A comparison of KFL learners' reading ability and vocabulary/grammar competence - focusing on the comparison between *Hanja* culture learners and non-*Hanja* culture learners. The Language and Culture 2017;13:1-22.
4. Honaker J, King G, Blackwell M. Amelia II: a program for missing data. J Stat Softw 2011;45:1-47.
5. Reynolds A, Richards G, de la Iglesia B, Rayward-Smith, V. Clustering rules: a comparison of partitioning and hierarchical clustering algorithms. J Math Model Algorithm 1992;5:475-504.
6. Bates D, Maechler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67:1-48.