An Exploration of High School Reading Test Items Based on Discriminant Analysis*

Nayu Kim·Heechul Lee**

(Chonbuk National University)

Kim, Nayu & Lee, Heechul. (2017). An Exploration of High School Reading Test Items Based on Discriminant Analysis. *The Linguistic Association of Korea Journal,* *25*(1), 75-92. The purpose of this study was to analyze dimensions of high school English reading test items based on discriminant analysis[1]. For this study, 387 high school students participated in the English reading test consisting of 28 items. The subjects' responses to the items were submitted to SPSS version 20 for discriminant analysis to investigate the dimensions of reading test items. The findings of the discriminant analysis revealed that the controlled items made by the examiners did not correctly discriminate the high level students. The categorization of reading test items which help divide the students into the high, intermediate and low level of whole subjects consisted of fourteen items, which were categorized as five types such as filling the blank, title inference, inserting sentence, completing the key sentence and text order inference. In addition, the discriminant function resulted from this study classified subjects as three groups according to their English ability correctly on upper 80% accuracy. The findings of this study indicated that specific types of reading test items played an important role in judging and predicting individual student' English proficiency. (Chonbuk National University)

Key Words: high school English reading test items, discriminant analysis, types of item

1) Discriminant analysis is the multivariate technique when the single dependent variable is dichotomous or multichotomous and therefore nonmetric.

1. Introduction

English reading test items played a significant role in assessing the students' English proficiency and categorizing them as low, intermediate or high level based on their scores. The high school students experienced various English reading test, which were composed of the mock version of scholastic aptitude test (SAT) English session and periodical formal achievement test controlled by the teachers. The former has been more frequent rather than the latter. Both of the teachers and students usually depended on the English mock test scores to identify their proficiency (Storey, 1997). Moreover, students selected and adopted their learning strategies and styles, and teachers tried to make various lesson plans according to the students' reading test scores based on standardized test format (Kintsch and van Dijk, 1978).

However, there were few studies about the effectiveness of individual reading test item to help find out the students' English proficiency. Particular types of reading test items are more powerful than the others on dividing the students into different levels (Grabe, 2001; Rumelhart, 1980). The examiners would view distinctive types of items as the controlled ones such as filling the blank and making inference, to classify the students as high or low level. Higher points are normally allotted to the controlled items. On the other hand, it is not asserted that each controlled item has equivalent value for identifying high level students (Block, 1986). Occasionally the other items other than the controlled ones have strength on it (Carrell, 1983; Yamashita, 2003).

Particularly, it is difficult to find out exertions involved in proving effectiveness of controlled items, which in advance evaluator scores many points to, even though there has been a few studies which analyzed the characteristics of item difficulty. In addition, the teacher as a evaluator has not investigated how to reflect the discriminant items on their lesson plan. With the main objective of identifying specific reading test items which discriminate the high school students as high, intermediate or low level on whole assessment context, the research questions were followed as below:

1) What types of English reading test items discriminate high school students as high, intermediate or low level without investigating whole items?

2) How closely are the controlled items connected to the discriminant items?

3) How do the teachers use the discriminant items on their designing lesson plans?

2. Theoretical Background

The researchers have investigated on English reading test items on multi-dimensions. Non-parametric dimensionality assessment procedures were used to evaluate the utility of reading test scores (Stout, 1987; Stout & Zhang, 1999; Ackerman & Shu 2009). They analyzed the effectiveness of individual reading test item on total scores. According to the results, the items which aimed principally at deduction rather than finding information played a role in subjects being higher ranker consistently. In addition, Hanson and Harris (1991) investigated validity of test items based on beta-binomial models and observed score distributions. The beta-binomial distribution was an array of binomial distribution which is frequently used to model the number of successes in a sample of size n drawn with replacement from a population of size N. They compared the result of evaluating subjects' performance to the efficacy of each item on the beta-binomial distribution, which revealed that analysis of reading test items based on the beta-binomial distribution was statistically different with the analysis of normal distribution. Moreover, factor analytic approaches were used for analyzing the reading test items statistically (Ling, 2009; Wainer et al., 2001; Stone et al., 2010). They extracted more powerful reading test items to have analogous characteristics and effect on total performance. Another approach to test reading items was to employ both a multidimensional and a unidimensional items response theory (IRT) model. von Davier (2008) investigated that individual examinee had noticeable response on specific reading test items, in which the examinee accustomed to thinking logically had higher grades on multiple choice items rather than one who tended to think by intuition. In addition to these investigations, Gibbons and Hedeker (1992) suggested the bi-factor analysis as a encouraging alternative to the previous works in analyzing reading test items (Haladyna & Kramer, 2004; Hays, Morizot & Reise, 2007; Hong, Song & de la Torre, 2011). The bi-factor approach was special version of full-information item factor analysis (FIFA), which was an

IRT-based factor analysis (Bock, Gibbons & Muraki, 1988). An advantage of using bi-factor analysis on test items was that one need not provide a specific bi-factor model a priori, which meant that each item was investigated randomly.

On the other hand, some researchers have analyzed the relation between specific types of items and the effectiveness on total test context. Han (2007) studied the types of a completion cloze and multiple choice cloze for skilled and less skilled undergraduate in Korea. The frequency and quality of the types of items were examined along with subjects' total test performance. The results showed that the more difficult the cloze test items were, the more the difference of total scores between two groups was. In addition, Choi (2010) analyzed the properties of reading test items based on IRT. He suggested that the type of filling the blank distinguished between low and high level subjects. On the same context, the examining each test items statistically through pilot test helped decide about selecting reading test items and contributed to validating reading tests (Spolsky, 1990; Kitao & Kitao, 1997; Bachman, 1998).

Even though there has been an array of studies related to reading test items, few of researchers have tried to employ the discriminant analysis on that field. Koo and Yang (2013) suggested the method of analyzing difficulty of text and readability based on discriminant analysis. The study investigated that specific factors was statistically significant on difficulty of text, and analyzed accuracy and influence of these factors on it. They suggested that four factors of grammatical unit, word difficulty, sentence length and adorning expression were statistically significant components. As the result, three factors except sentence length were meaningful and two discriminant functions were elicited. The first function confirmed approximately 98.5% of group discriminant, and difficulty of word showed the strongest influence. The second function had an explain 1.5% and the adorning expression factor showed bigger influence than other factors.

3. Research Method

3.1 Subject

In Chonbuk province, 387 high school students who was in same high school

participated in this research. The results of reading (28 items) and listening (17 items) proficiency tests (total 45 items) conducted before one month showed that average English level was about 4.5 out of 9, which meant intermediate level. All of subjects were males and had not been more than one semester in English-speaking countries. The subjects were in the same curriculum, in which students had five English classes a week and did not have any classes with native speaker.

Table 1. Subjects on Their English Ability

		English level	Students(N)	%
Low ↑ ↓ High	1~3	Low group	6	1.6
			12	3.1
			58	15.0
	4~6	Intermediate group	87	22.5
			72	18.6
			54	14.0
	7~9	High group	52	13.4
			20	5.2
			7	6.7
	Total		387	100.0

For the purpose of discriminant analysis, the subjects were classified as 9 groups based on their total test scores. As a result, the high level group from level 7 to 9 consisted of 98 students, while intermediate level, from level 4 to 6, 213 students and low level, from level 1 to 3, 76 students. The specific number of students on the individual group and ratio were as in Table 1.

3.2. Instrument

The reading test items investigated in this study were conducted in classrooms in the school on April 2013. The subjects of three groups were asked to respond to 28 reading test items, all of which were the selection type consisting of an answer and 4 distractors. The items were classified into 15 types based on categorizations of reading test items of College Scholastic Ability Test.

The test items were created by the Public Office of Education. The specific information of the items was as in Table 2.

Table 2. Items Information

No.	Item classification	Scores (controlled)
18	finding the objective	2
19	identifying the reference	2
20	title inference	2
21	gist inference	2
22	main idea inference	3
23-24	vocabulary inference	2-3
25	grammar	3
26	finding the specific information	2
27	inserting the sentence	2
28	finding non-related sentence	2
29	analyzing the diagram	2
30	text order inference	2
31	finding the specific information	2
32-34	filling the blank	2
35-37	filling the blank	3
38-39	filling the blank	2
40	completing the summarizing	3
41	text order inference	2
42	identifying the reference	2
43	finding the specific information	2
44	title inference	2
45	filling the blank	2

The test items were divided into two group, which were 2 points items and 3 points items, where the latter was considered as more difficult to the students than the former. The items were categorized to fifteen types, from finding the objective to filling the blank. The subjects answered each item per 90 seconds since time limit was 50 minutes on the whole test. The test was conducted in

whole classes in the high school at the same time and the teachers who worked in the school controlled the test procedure.

3.3 Data Analysis

For analyzing the dimensions of reading test items, discriminant analysis was conducted. In this study, the dependent variables was the students' English proficiency level which was the ordinal variables; low level group (level 1 to level 3), intermediate group (level 4 to 6) and high level group (level 7 to 9). The independent variables were the students' responses to the test items in which their correct answering was coded for '1' and incorrect answering was coded for '0'. To obtain the validity of the analysis, 'stepwise' method was adopted. To analyze the probability of the discriminant analysis, 'Box's M test' was conducted, and 'Wilk's Lambda', 'Eigenvalues & Level of Statistical Significance' and 'Canonical Discriminant Function Coefficients' was investigated. In addition, 'Classification Results' and 'Canonical Discriminant Functions' were followed.

4. Results & Discussion

4.1 Key Items of Discrimination

Each item had individual influences according to the students' level. In low level group, the highest mean value was .75, (item 31) and the lowest one was .04 (item 35). In intermediate level group, the highest mean value was .95 (item 31) and the lowest one was .06 (item 44). Lastly, in high level group, the highest mean value was .99 (item 26) and the lowest one was .32 (item 35). The highest and the lowest mean value gradually increased as the level of students increased as in Table 3.

Table 3. Mean & S.D of English Ability Groups on Items

Item	Low level group		Intermediate level group		High level group	
	Mean	S.D	Mean	S.D	Mean	S.D
18	.67	.473	.93	.256	.96	.199
19	.08	.271	.46	.500	.90	.304
20	.24	.428	.41	.494	.78	.419
21	.29	.457	.79	.406	.98	.142
22	.32	.468	.65	.477	.90	.304
23	.25	.436	.57	.496	.93	.259
24	.18	.390	.53	.501	.76	.432
25	.17	.379	.23	.422	.57	.497
26	.58	.497	.88	.328	.99	.101
27	.14	.354	.22	.416	.68	.467
28	.33	.473	.66	.476	.94	.241
29	.38	.489	.74	.441	.96	.199
30	.32	.468	.53	.500	.96	.199
31	.75	.436	.95	.222	.98	.142
32	.22	.419	.45	.498	.83	.381
33	.16	.367	.45	.499	.84	.372
34	.20	.401	.27	.446	.57	.497
35	.04	.196	.08	.279	.32	.467
36	.09	.291	.19	.391	.49	.502
37	.08	.271	.26	.441	.47	.502
38	.26	.443	.24	.431	.48	.502
39	.11	.309	.13	.333	.49	.502
40	.08	.271	.12	.328	.33	.471
41	.09	.291	.29	.453	.76	.432
42	.28	.450	.24	.431	.54	.501
43	.47	.503	.58	.494	.90	.304
44	.09	.291	.06	.231	.37	.485
45	.11	.309	.22	.412	.48	.502

At each step, the variable that minimized the overall Wilks' Lambda was entered. Fourteen variables were entered according to F values and fourteen items were removed as in Table 5. The F value of item 18, 19, 21, 23, 24, 25, 27, 30, 32, 33, 37, 39, 41 and 44 were statistically significant and their significant levels were .001, which meant that the discriminant functions had 14 variables. F value of the individual item was from 32.176 to 83.523. Wilks' Lambda was .204 until 14th variable was entered, which meant the presented analysis was

statistically significant (R^2=1-λ). The highest increase in Wilks' Lambda was .697 to item 19 and the lowest one was .006 (.210-.204) from item 37 to item 27 as in Table 4.

Table 4. Wilks' Lambda & F of Items

Step	Entered	Wilks' Lambda	F	Sig.	df3
1	19	.697	83.523	.001	384
2	21	.539	69.443	.001	384
3	41	.444	63.813	.001	384
4	33	.371	61.107	.001	384
5	39	.338	54.703	.001	384
6	18	.312	50.005	.001	384
7	32	.289	46.505	.001	384
8	23	.270	43.599	.001	384
9	44	.254	41.158	.001	384
10	25	.240	39.021	.001	384
11	24	.229	37.001	.001	384
12	30	.220	35.262	.001	384
13	37	.210	33.873	.001	384
14	27	.204	32.176	.001	384

As seen in Table 5, as a result of the discriminant analysis of 14 meaningful test items, two discriminant functions extracted. In order to investigate the validity statistically of two functions, the Chi-square test was conducted, all of which was statistically significant (p<.05). Function 1 and Function 2 were all significant, of which the Function 1 explained 90.3% of total variance and the Function 2 explained 9.7% of total variance. The Function 1 was more statistically significant than the Function 2 but the latter had enough portion on the analysis. In addition, the canonical correlation on Function 1 was .858 and one on Function 2 was .479, which revealed the significance of Function 1.

Table 5. Eigenvalues & Level of Statistical Significance

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation	Chi-square	df	Sig.
1	2.780	90.3	90.3	.858	600.143	28	.001
2	.297	9.7	100.0	.479	98.215	13	.001

To investigate the characteristics of Function 1 and Function 2, the coefficients of variables of two discriminant functions were analyzed on as in Table 6. All fourteen variables had influence in Function 1, where item 19 (.983) and item 33 (.906) were the highest coefficients[1] among them. Otherwise, item 27(.269) and item 44(.353) were the lowest ones in Function 1. In Function 1, relatively item 19 (.416) had the highest influence and item 44 (.115) had the lowest influence.

Table 6. Canonical Discriminant Function Coefficients

item	Unstandardized Function		Standardized Function	
	1	2	1	2
18	.540	-1.329	.162	-.399
19	.983	-.209	.412	-.088
21	.791	-1.570	.292	-.581
23	.641	-.156	.279	-.068
24	.470	-.473	.218	-.220
25	.490	.543	.213	.236
27	.269	.702	.113	.294
30	.507	.369	.222	.161
32	.672	.181	.306	.082
33	.906	-.147	.403	-.066
37	.555	-.212	.239	-.091
39	.523	.821	.198	.311
41	.728	.406	.306	.171
44	.353	1.255	.115	.407
constant	-4.083	1.709		

Likewise, all fourteen variables had influence in Function 2 as well where item 44 (1.255) was the highest coefficients among them. However, item 18 (-1.329) and item 21 (-1.570) were the lowest ones in Function 2. In Function 2, relatively item 44 (.407) had the highest positive influence and item 18 (-.399) had the lowest influence.

Function 1 and Function 2 were constructed as followed.

---

1) In the function as 'y = ax^2 + bx + c,' it is generally supposed that x is the only variable and that a, b and c are coefficient; thus the constant coefficient is c in this case.

Function 1 = -4.083 + .540(item 18) + .983(item 19) + .791(item 21) + .641(item 23) + .470(item 24) ...... + .353(item 44)

Function 2 = 1.709 + -1.329(item 18) + -.209(item 19) + -1.570(item 21) + -.156(item 23) + -.473(item 24) ...... + 1.255(item 44)

The centroids[2] of each group classified by 14 items were extracted as in Table 7. The centroid of low level group was − 2.392, intermediate level group -.298, the high level group 2.502, which meant that when the value of 〔-4.083 + .540(item 18) + .983(item 19) + .791(item 21) + .641(item 23) + .470(item 24) ⋯ + .353(item 44)〕 of any student's response on the test were close to any presented centroids, the student could be discriminated as the specific group member.
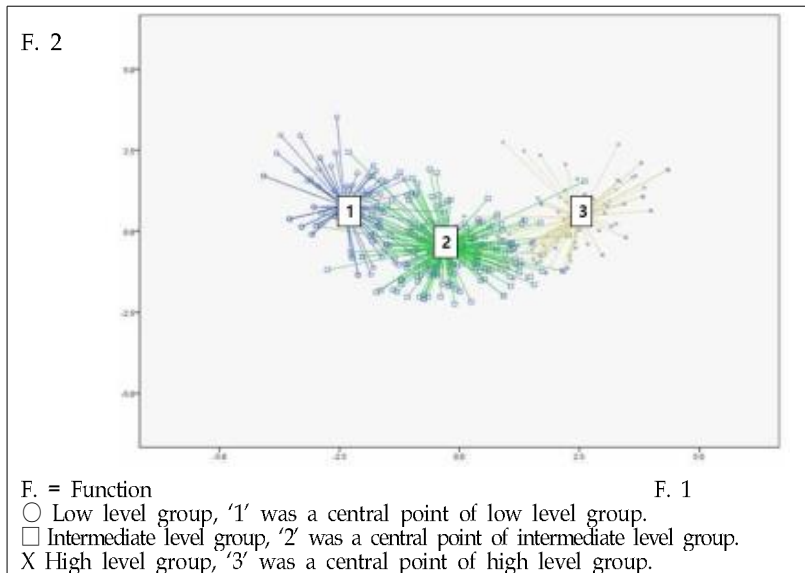
Table 7. Functions at Group Centroids

Group	Function 1	Function 2
Low level	-2.392	.771
Intermediate level	-.298	-.481
High level	2.502	.448

When Function 1 was a horizontal axis and Function 2 vertical axis, discriminant functions of each group were mapped as in Figure 1. The characteristics of low, intermediate and high level group were correctly discriminated by Function 1.

---

2) Mean value for the discriminant Z scores of all objects within a particular category or group. For example, a two-group discriminant analysis has two centroids, one for the objects in each of the two groups.

Figure 1. Canonical Discriminant Functions[3)]

F. = Function F. 1
○ Low level group, '1' was a central point of low level group.
□ Intermediate level group, '2' was a central point of intermediate level group.
X High level group, '3' was a central point of high level group.

4.2 Validity of Controlled Items

Based on two discriminant functions, the coefficients of discriminant functions of each group were analyzed as in Table 8. As seen Table 8, all fourteen variables had influence in low level group, where item 18 (7.389) was the highest coefficients among them. Otherwise, item 19 (.332) and item 41 (.371) were the lowest ones in low level group. Likewise, in intermediate group, item 18 (10.184) was the highest coefficients among them. However, item 44 (-.106) was the lowest ones in intermediate group. Lastly item 18 (10.461) was the highest coefficients but item 27(1.523) was the lowest ones in high level group.

---

3) A variate of the independent variables selected for their discriminatory power used in prediction of group membership.

Table 8. Classification Function Coefficients

Item	Group		
	Low level	Intermediate level	High level
18[*]	**7.389**	**10.184**	**10.461**
19[*]	.332	**2.653**	**5.211**
21[*]	**2.056**	**5.678**	**6.434**
23[*]	**1.107**	**2.644**	4.293
24[**]	.689	2.265	3.141
25[**]	.628	.975	2.852
27[*]	.432	.117	1.523
30[*]	1.046	1.647	3.409
32[*]	**1.141**	2.323	**4.371**
33[*]	**1.090**	**3.171**	**5.570**
37[**]	.641	2.069	3.426
39[*]	.862	.929	3.157
41[*]	.371	1.387	3.802
44[*]	.725	-.106	2.048
(Constant)	-4.675	-12.369	-25.282

* 2 points item
** 3 points item

Each discriminant function based on each group was constructed as followed. When the student's response of each variable was assigned on each function, he or she was classified as the group whose function had the highest total estimates among three functions.

Function of 'Low level group' = -4.675 + 7.389(item 18) + .332(item 19) + 2.056(item 21) + 1.107(item 23) + .689(item 24) ...... + .675(item 44)

Function of 'Intermediate level group' = -12.369 + 10.184(item 18) + 2.653(item 19) + 5.678(item 21) + 2.644(item 23) + 2.265(item 24) ...... + -.106(item 44)

Function of 'High level group' = -25.282 + 10.461(item 18) + 5.211(item 19) + 6.434(item 21) + 4.293(item 23) + 3.141(item 24) ...... + 2.048(item 44)

Particularly among seven 3 points items originally controlled by the official examiner, none of them were in the five highest coefficients of 'Function of Low

level group' and 'Function of High level group,' which meant those items did not play an important role in discriminating the subjects on their English proficiency. However, item 18, item 21 and item 33 had commonly significant influence in discriminating the students as three groups.

Based on the discriminant analysis on this study, the classification results of 387 students were as in Table 9. In low level group, 64 students of the original group were exactly classified as the predicted group. In intermediate level group, 171 students of the original group were exactly classified as the predicted group. In high level group, 87 students of the original group were exactly classified as the predicted group. Accordingly, 83.2% of original grouped students correctly classified. Therefore 14 variables had significant influence on discriminating the students on three level groups.

Table 9. Classification Results

		Group	Predicted Group Membership			Total
			Low	Intermediate	High	
Original	Count	Low	**64**	12	0	76
		Intermediate	33	**171**	9	213
		High	0	11	**87**	98
	%	Low	**84.2**	15.8	.0	100.0
		Intermediate	15.5	**80.3**	4.2	100.0
		High	.0	11.2	**88.8**	100.0
Cross-validated	Count	Low	64	12	0	76
		Intermediate	34	**169**	10	213
		High	0	11	87	98
	%	Low	**84.2**	15.8	.0	100.0
		Intermediate	16.0	**79.3**	4.7	100.0
		High	.0	11.2	**88.8**	100.0

5. Conclusion

In order to identify specific reading test items which discriminate the students according to their English level on whole assessment context, the discriminant functions were extracted, which had statistically significant prediction. The findings of this study revealed that the high school English

reading test items were investigated by the discriminant analysis and through the analysis, some kinds of items had more superior status relative to the other types of items, compared to the study of Koo and Yang (2013) which focused on the difficulty and readability of the text. In particular, the types of selecting vocabularies and using grammar which the teachers considered as the decisive items in fact did not have high effectiveness distinguishing the high-level from the low level students.

The findings of the present study provided important implications on analyzing the reading test items in the same veins of the study of Choi (2010) based on IRT. First, the controlled items given higher scores did not play a significant role in discriminating the students in assessing procedures. As seen in Table 8, the students responded to the items based on the five specific types such as filling the blank, title inference, inserting sentence, completing key sentence and text order inference. In this context, the effect of the vocabularies and grammar test items was considerably limited. Among seven 3 points items, just three items such as item 35, 36 and 40 could effect on discriminating the students.

Second, teachers can create filling the blank, title inference, inserting sentence, completing key sentence and text order inference item as a reference to verify students' level when they designed the lesson plan. In other words, teachers can devise simple items to identify students' level, no need of whole type of reading items. In addition, They can apply key items to checking students complete their performance by periodical formative test.

However, future studies with more varied research method will be needed in that the present study focused on high school reading test items. In particular, the subject and the high school of this research are not most representative of Korean high school and its students. In addition, reading and listening test items in middle school can be analyzed based on a multitude respect including multi-variate statistics, and listening test items in high school can be investigated on various dimensions since there are characteristics of test items involved in subject's developmental stage and difference between reading and listening test items.

References

Ackerman, T., & Shu, Z. (2009). Using confirmatory MIRT modeling to provide diagnostic information in large scale assessment. Paper presented at the annual meeting of the National Council of Measurement in Education, San Diego, CA.

Bachman, L. F. (1998). Language testing-SLA research interfaces. In L. F. Bachman & A. D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 177-195). Cambridge: Cambridge University Press.

Block, E. (1986). The comprehension strategies of second language learners. *TESOL Quarterly, 20*, 463-494.

Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement, 12*, 261-80.

Carrell, P. L. (1983). Three components of background knowledge in reading comprehension. *Language Learning, 33*(2), 183-207.

Choi, K.-H., & Lee, J.-H. (2010). Devising of proper items for the evaluation of English proficiency test: A case study in J university. *Modern Studies in English Language & Literature, 54*(3), 171-186.

von Davier, M. (2008) A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology, 61*, 287-307.

van Dijk, T., & Kintsch, W. (1978). Towards a model of text comprehension and production. *Psychological Review, 85*, 363-394.

Gibbons, R. D., & Hedeker, D. R. (1992). Full-information item bi-factor analysis. *Psychomerika, 57*, 423-36.

Grabe, W. (2001). *Reading-writing relations: Theoretical perspectives and instructional practices*. In D. Lcher & A. Hirvelam (Eds.), Linking literacies: Perspectives on L2 reading-writing connections (pp. 15-47). Ann Arbor, MI: Michigan University Press.

Haladyna, T. M., & Kramer, G. A. (2004). The validity of subscores for a credentialing test. *Evaluation and the Health Professions, 27*(4), 349-68.

Han, Mihyang. (2007). EFL readers' test-taking processes for completion vs. multiple choice cloze tests. *The Linguistic Association of Korea Journal, 15*(3), 189-208.

Hanson, B. A., & Harris, D. J. (1991). *Methods of examining the usefulness of subscores.* Paper presented at the meeting of the National Council on Measurement in Education, Chicago, IL.

Hays, R. D., Morizot, J., & Reise, S. P. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research, 16,* 19-31.

Hong, Y., Song, H., & de la Torre, J. (2011). A comparison of four methods of IRT subscoring. *Applied Psychological Measurement, 35*(4), 296-316.

Kitao, S. K., & Kitao, K (1997). Validity and reliability, in TESL-L Electric Discussion Forum.

Koo, M., & Yang, K. (2013). Discriminant Analysis of Korean Reading Text Difficulty. *Bilingual Research, 52,* 1-18.

Lane, S., Stone, C. A., Ye, F., & Zhu, X. (2010). Providing subscale scores for diagnostic information: A case study when the test is essentially unidimensional. *Applied Measurement in Education, 23,* 63-86.

Ling, G. (2009). *Why the major field test does not report subscores of individual test-takers −Reliability and construct validity evidence.* Paper presented at the annual meeting of American Educational Research Association and the National Council on Measurement in Education.

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce & W. E. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 33-58). Hillsdale, NJ: Lawrence Erlbaum.

Spolsky, B. (1990). Social aspects of individual assessment. In J. H. A. de Jong & D. K. Stevenson. *Individualizing the assessment of language abilities.* Clevdon, Avon, England: Multilingual Matters Ltd.

Storey, P. (1997). Examining the test-taking process: A cognitive perspective on the discourse cloze test. *Language Testing, 14*(2), 214-231.

Stout, W. (1987). A non-parametric approach for assessing latent trait unidimensionality. *Psycholmetrika, 52,* 589-617.

Stout, W., & Zhang, J. (1999). The theoretical DETECT index of dimensionality and its application to approximate simple structure. *Psychometrika, 64,* 213-49.

Wainer, H, Vevea, J. L., Camacho, R., Reeve, B. B., Swygart, K. A., & Thissen, D. (2001). Augmented scores-borrowing strength to compute scores based

on small numbers of items. In D. Thissen & H. Wainer (Eds), *Test Scoring*. Mahwah, NJ: Lawrence Erlbaum.

Yamashita, J. (2003). Processes of taking a gap-filling test: Comparison of skilled and less skilled EFL readers. *Language Testing, 20*(3), 267-293.

Kim, Nayu
Department of English, Wonkwang highschool
9-4, Busongdong, Iksan, Chonbuk Province
Phone: 063-723-3654 / C.P: 010-3689-1362
E-mail: kimnayu98@naver.com

Lee, Heechul
Department of English Education, Chonbuk National University
664-1 Deokjingu, Jeonju, Chonbuk Province
Phone: 063-270-2719
E-mail: hclee@cbnu.ac.kr