Volume 14(1), February 2014 (http://dx.doi.org/10.17576/GEMA-2014-1401-06)

Investigating Chinese Language Learners' Reading Comprehension for Different Meaning Types

Mi Soon Kim
misoon@hufs.ac.kr
Hankuk University of Foreign Studies,
Republic of Korea

Hyun Jung Kim (corresponding author)
hkim@hufs.ac.kr
Hankuk University of Foreign Studies,
Republic of Korea

ABSTRACT

The Chinese government has published an official guide specifying the aims for reading in Chinese and the expected comprehension levels for different proficiency learners, with regard to teaching and learning Chinese as a second or foreign language. However, due to lack of teacher training for its implementation, this guide has rarely been used for teaching reading to Chinese language learners and has rarely been used for evaluating their reading ability. Therefore, it appears that teaching and assessing reading comprehension have not been based on a theoretical background of reading ability. The purpose of this study is twofold: (1) to provide validity evidence of a Chinese reading test developed based on a theoretical model of reading ability; and (2) to examine reading test performance of Chinese learners with various reading ability levels. For the purpose of this study, reading ability was defined based on a meaning-based model that included three layers of reading comprehension: literal, intended, and implied meanings of a reading text. A total of 248 Korean university students were divided into three levels, and their test performances were analyzed and compared for the three meaning types using structural equation modeling and regression analysis. The results suggest that the test performance structure represented the meaning-based model in general, thereby providing validity evidence of the test. Further analyses revealed that the three groups differed from one another with respect to their understanding of literal, intended, and implied meanings. The findings provide pedagogical implications for teaching Chinese language learners with different reading proficiency levels.

Keywords: Chinese as a second/foreign language; second language reading ability; reading test; New HSK; reading item type

INTRODUCTION

The emergence of Communicative Language Teaching (CLT) and theoretical models of communicative competence have influenced not only *teaching* second or foreign languages, but also *assessing* language ability. Teachers have tried to focus more on fluency, purposeful/meaningful communication, and the use of authentic materials in language classrooms, rather than on the accurate usage of language, pattern drills, and practice. Emphasis on communicative language use has also brought about gradual changes in defining test constructs. Recently, constructs of reading ability have been defined for the evaluation of language learners' ability to understand the surface and underlying meaning of a writer's message, instead of the evaluation of learners' ability to understand sentence structures or sentence-level meanings.

ISSN: 1675-8021

Such a highlight on the communicative aspects in reading is also evident in the language education policy for Chinese language learners. The Chinese Language Proficiency Scales for Speakers of Other Languages, developed by the Office of Chinese Language Council International (2007), serve as criteria for Chinese teaching, learning and assessment, focusing on real-world language use. According to the scales, reading comprehension in Chinese is defined as including the comprehension of (i) correspondence in social interactions (e.g., a card of congratulations from a friend and private correspondence), (ii) instructive and explanatory texts (e.g., brief introductions to new books and popular science articles), and (iii) various kinds of informative texts (e.g., posters on a college campus and job advertisements). Such definitions reflect a concentration not only on literal meaning, but also on the pragmatic aspects of the Chinese language, unlike the Chinese language teaching and learning of the past (Zhou & Li, 2009).

The scales further define Chinese reading ability separately for different ability levels. The beginner level is defined as the ability to read and understand simple narrative or descriptive texts that are related to everyday life, grasping the main and concrete ideas. The intermediate level further includes the ability to understand the intention of the author, in addition to the requirements of the beginner level. The advanced level is defined as the ability to understand abstract, conceptual or technical information from the texts, read between the lines, and understand the author's viewpoints or intentions. As such, different purposes of reading education and assessment have been established, with alternating lengths and types of reading texts and levels of meaning required for comprehension.

Even though reading ability has been described concretely in the scales, the definitions of different levels of reading ability have rarely been operationalized for reading assessment in China, as well as in other countries teaching Chinese as a second or foreign language, whether it be a high-stakes large-scale test or a classroom assessment. In addition, there seems to be little research that seeks to validate these reading tests (e.g., Kim & Park, 2013; Jeong, 2008; Seong, 2010) based on a theoretical model of reading ability (e.g., the Chinese Language Proficiency Scales for Speakers of Other Languages).

To this end, the present study aimed to (1) define the construct of Chinese reading ability based on a theoretical framework, and (2) investigate the extent to which Chinese learners of various proficiency levels understand different layers of meanings (i.e., literal, intended, and implied). The findings of this research are expected to suggest theoretically defined, and empirically operationalized reading ability for Chinese teaching and learning. Ultimately, a deeper understanding of reading ability will lead test developers and teachers not only to develop a reading test measuring learners' understanding of various layers/types of meanings beyond the literal meanings of texts, but will also lead them to make appropriate interpretations of learners' reading ability.

LITERATURE REVIEW

DEFINING AND TESTING SECOND LANGUAGE (L2) READING ABILITY

There have been continuous attempts to define reading ability in L2 reading research. Traditionally, L2 reading ability has been defined in terms of its processes or products/components. Many researchers who focus on reading processes try to depict the mental activities involved in reading. Such mental activities are most commonly discussed using three processing models: top-down, bottom-up, and interactive models. The top-down model emphasizes the importance of activating existing schemata and the involvement of readers' knowledge in the reading process (Alderson, 2000). Readers bring meaning to the text based on their prior knowledge and experience, and are actively involved in creating meaning out of the text. Contrary to the top-down model, which emphasizes readers' active

role, the bottom-up model considers readers as passive decoders who process independent, sequential graphic, phonemic, syntactic, and semantic systems in order. According to Alderson (2000), bottom-up approaches are "serial models, where the reader begins with the printed word, recognizes graphic stimuli, decodes them to sound, recognizes words and decodes meanings" (p. 16). Top-down and bottom-up models explain the reading process in different ways; however, neither processing model alone can fully explain the reading process. When individuals are engaged in reading, they selectively employ either the top-down or bottom-up process to comprehend meanings and compensate for deficiencies in the other. According to the interactive model, top-down and bottom-up approaches can occur either at the same time or alternately, depending on the reading texts, reader characteristics, and reading purposes (Alderson, 2000).

The other attempt to define reading ability focuses on its products or components. Since reading ability is regarded as divisible, many researchers have tried to identify the separate elements involved in reading. This view has not only encouraged researchers to propose taxonomies of reading skills to define what it means to be able to read (e.g., Carroll, 1993; Grabe, 1991), but it has also provided implications for L2 reading assessment. Weir (1997) argues that "if specific skills, components or strategies could be clearly identified as making an important contribution to the reading process, then it would of course be at least possible […] to test these and to use the composite results for reporting on the reading proficiency" (p. 44). Oftentimes, sub-skills identified under reading ability (e.g., identifying the main idea, understanding details, and inferencing) have served as operational definitions of reading ability, especially for large-scale multiple-choice tests (e.g., Alderson, 1990; Lumley, 1993).

Since neither attempt can fully explain reading performance without the other, researchers have recently proposed a new approach to define reading ability, incorporating both the process and product of reading. Within a broader framework of language knowledge (Purpura, 2004), Liao (2008) and Kim, A.Y. (2011) define reading ability in terms of the different types of meanings obtained from the text. Purpura's (2004) language knowledge is divided into (1) grammatical knowledge, including grammatical form (various linguistic forms) and grammatical meaning (literal and intended meaning of utterances), and (2) pragmatic knowledge, including contextual, sociolinguistic, sociocultural, psychological, and rhetorical meanings. While explaining language knowledge, Purpura (2004) differentiates three different types of meanings involved in language ability: literal, intended, and pragmatic/implied meanings. Kim, A.Y. (2011) adopted Purpura's (2004) meaning-based framework and operationalized L2 reading ability in terms of three types of meanings (literal, intended, and implied). That is, literal meaning, which requires understanding the surface-level meaning derived from a text, is differentiated from the other two types of intended and pragmatic meanings, which require some sort of inferencing. Intended and pragmatic meanings are distinguished based on whether inferencing is made within the text (intended) or outside/beyond the text (pragmatic/implied). Kim, A.Y. (2011) argues that this new approach to define reading ability better explains L2 reading than earlier attempts, since the primary purpose of reading is to understand meanings that are derived in various ways from texts. Therefore, the three types of meanings are essential in understanding and testing individuals' reading ability.

However, there have been only few attempts to define and test reading ability based on such a meaning-based model, especially in a second or foreign language other than English. Specifically, in the case of Chinese reading assessment, the Chinese Language Proficiency Scales (2007) define reading ability by focusing on different types of meanings that learners are expected to understand at different proficiency levels. Even though the definitions are very similar to the three types of meanings specified in Purpura's (2004)

model, it is unclear whether or not these definitions came from any theoretical framework of reading ability or proficiency. Moreover, reading ability has not been operationalized for Chinese reading test development. Therefore, the present study aims to provide empirical evidence of the construct validity of a Chinese as a foreign language (CFL) reading test, which has been developed based on Purpura's (2004) meaning-based model. In addition, this study aims to explore the nature of CFL students' reading ability at different levels by investigating the extent to which they understand different types of meanings while reading.

TESTING THE READING ABILITY OF CFL LEARNERS

For teaching and learning Chinese reading, extensive (top-down model) and intensive (bottom-up model) reading approaches have widely been used (e.g., Chen, 2010; Zhao, 2004; Zhou & Li, 2009). As Chinese language teaching at the secondary level puts equal weight on all four skills of listening, speaking, reading and writing and emphasizes the integration of language skills, integrated-skills tests are the most common type of assessment. On the other hand, in higher education the four skills are often taught separately as independent courses (Zhou & Li, 2009). Therefore, Chinese language courses at the level of higher education often use independent-skills tests (e.g., achievement and diagnostic tests) specified for each literacy area (e.g., reading). However, it seems at present that there is very little research on how to test independent language components or skills at the higher education level.

As the New Hanyu Shuiping Kaoshi (New HSK) is the most common and representative Chinese language test, most scholarly discussions on Chinese language assessment focus on assessment using the New HSK (e.g., Kim A.Y., 2011; Jeong, 2008; Seong, 2010). The New HSK is the most well-known Chinese language proficiency test developed based on the Chinese Language Proficiency Scales for Speakers of Other Languages, and is utilized worldwide. While it is the most renowned test, most research on the New HSK seems to concentrate on the test format rather than on the test construct. This is also the case for research on the reading comprehension tests under the New HSK. For example, previous research has addressed such issues as evaluation of the testing system (Jeong, 2008; Seong, 2010) and comparisons between the New HSK and other Chinese language tests (Kim M. S., 2011). It is difficult, however, to find research on test constructs or post-assessment feedback. In particular, studies on testing reading ability based on the meaning-based model are not yet to be seen. Therefore, research that involves defining the test construct of reading ability and then conducting an assessment is deemed to have significant academic value.

Because the New HSK test items are not accessible to the public, most studies on the New HSK (Kim M. S., 2011; Jeong, 2008; Seong, 2010) have been based on the Chinese Language Proficiency Test Syllabus Levels 1 to 6 (The Office of Chinese Language Council International & Confucius Institute Headquarters, 2009a, 2009b, 2009c, 2009d, 2009e, 2009f), which provide an overview of the content of the New HSK. According to the syllabus, test items in the reading comprehension section include a broad array of topics such as daily life, politics, economics, society, and culture. However, the questions center mostly on meaning, and in particular, literal meaning rather than on intended meaning. In addition, there are not many pragmatics-related items, thereby resulting in a limited number of items on implied meaning (Kim & Park, 2013). Such being the situation, the test construct was defined based on the meaning-based model, and test items were newly developed to include all three types of meanings, while reading texts were directly cited from the Chinese Language Proficiency Test Syllabus.

RESEARCH PURPOSE

To reiterate, the present research aimed to examine the extent to which Chinese language learners of different reading ability levels can understand a variety of meanings through a reading test which was developed based on the operationalization of a construct definition of reading ability (Purpura, 2004). The research also investigated the differences in test performance based on different levels of reading ability. The study sought to answer the following research questions:

1. What is the underlying trait structure of foreign language test performance as measured by the Chinese reading test?
2. Do the three groups of CFL learners (beginner, intermediate, and advanced) exhibit differences in their scores on the components of reading ability?

METHODOLOGY

PARTICIPANTS

The study was carried out with 248 Korean university students as participants. All participants were attending a Chinese language course at a university located in Seoul. Initially, college students who had Chinese-related majors (e.g., Chinese linguistics and literature, Chinese regional studies, and Chinese translation and interpretation) were invited to volunteer to participate in the current study. Among the students who agreed to participate, only students speaking Korean as their first language were selected, and native Chinese-speaking students were not included. The participants ranged from first year to fourth year at the university, and thus reflected multiple levels of Chinese language proficiency. Consequently, the participants included 91 students at the beginning reading level, 110 at the intermediate level, and 47 at the advanced level. (More information about group classification and a detailed explanation of such reading levels are provided in the Procedures section.)

Five Chinese language experts participated in setting the criteria to determine the threshold scores that distinguished the students' reading levels (beginning, intermediate and advanced). These experts included two current Chinese language adjunct instructors, one HSK instructor, one doctoral candidate specializing in Chinese language education, and one university faculty member with a doctoral degree in Chinese language education. All five experts were females in their 30s to 50s. They had two to ten years of experience in teaching Chinese (e.g., conversation, HSK preparation courses, and other test preparation courses).

INSTRUMENTS

READING TEXT

The reading texts were taken from texts found in the Chinese Language Proficiency Test Syllabus. The New HSK is comprised of six levels, with the lowest being level 1. The basic levels of 1 to 3 were excluded from the range to be covered in the present test, while level 4 for beginners, level 5 for intermediate learners, and level 6 for advanced learners were all included. The basic levels were excluded because they focus on vocabulary and grammar instead of reading itself, considering low-beginner learners' limited language ability. For the reading texts, two short texts (1-2 sentences) and two lengthy texts (1 paragraph) were chosen from the beginning level 4, two short texts (2-4 sentences) and two lengthy texts (2-3 paragraphs) were chosen from the intermediate level 5, and two lengthy texts (3-4 paragraphs) were chosen from the advanced level 6. The reading texts included a variety of topics, such as finding friendship in hardship, the path to a good life, balancing money and time, and a great man's wisdom. The length of the text and the difficulty of the vocabulary

and grammar were considered in selecting the excerpts. One of the reading texts was shortened in order to develop test items using the deleted part of the text. The texts and accompanying test items are presented in the Appendix.

READING ITEMS

The construct definition of reading ability drew on Kim (2011) L2 reading ability, adapted from Purpura (2004)'s language ability model. That is, the assessment goal for the reading items was set at understanding literal, intended, and implied meanings. A total of 20 items were developed based on such assessment goals. The number of items was determined for practicality reasons, and the items were pilot tested in advance. All 20 questions were coupled with discrete-point multiple-choice items.

The test was composed of a total of 20 items, including seven on understanding literal meaning, eight on understanding intended meaning, and five on understanding implied meaning. Specifically, the test included six beginner-level items (three literal and three intended meaning questions), seven intermediate-level items (two literal, three intended, and two implied meaning questions), and seven advanced-level items (two literal, two intended, and three implied meaning questions). The item types are represented in Table 1. Examples of test items include "*Which of the following is correct according to the passage?*" (literal meaning); "*What is the author's intention?*" (intended meaning); and "*What phrase is most appropriate in the blank*?" (implied meaning). For the sake of designing questions based on the test construct, 16 out of the 20 items cited questions from the New HSK syllabus, while the other four items were custom designed by the researchers. The syllabus rarely included intended and implied meaning items; as a result, some items had to be newly developed in order to ask all three types of meaning on the test.

TABLE 1. Item types by proficiency level

Item type	Beginner	Intermediate	Advanced	Total
Literal meaning	3	2	2	7
Intended meaning	3	3	2	8
Implied meaning	–	2	3	5
Total	6	7	7	20

PROCEDURES

Data collection occurred in four stages: (1) The cut-off point for the score was established for reading ability level distinction; (2) The reading test was conducted; (3) The participants' responses were scored; and (4) Reading ability levels were identified based on test scores.

In the first stage, the Angoff procedure (Angoff, 1971) was designed for the multiple-choice reading test, which is used in the testing process to systematically distinguish test-takers' achievement/performance levels. It was used to determine the cut-off point for each level (beginning, intermediate, and advanced). A cut-off point to distinguish between the beginning and intermediate levels was established as follows. First, the five experts met to assess their understanding of each level. For each of the 20 items on the test, each member wrote down the prospective probability that students with a minimal competence level in the intermediate group would correctly answer each item. Then the five members compared one

another's numbers reflecting the probability in order to identify the items that demonstrated the biggest difference. Referring to opinions collected, probabilities were written down again. Then the probability scores were added up and divided by 20, which is the total number of test questions. Once an average was produced after collecting the results from all five members, the cut score between the beginning and intermediate levels was established. Following the same procedures, the probability for the minimal level students in the advanced level to correctly answer each item was determined in order to establish the cut score distinguishing between the intermediate and advanced levels. As a result of the Angoff procedure, the cut-off point between the beginning and intermediate levels was a score of 11 (out of 20), and the cut-off point between the intermediate and advanced levels was a score of 16 (out of 20).

In the second stage, the 248 participants took the reading test. The students were selected by convenience sampling. The participants who were asked to volunteer to participate in the present study were taking Chinese language courses at a university where one of the Chinese language experts (university faculty member) was teaching. The test was conducted in the 11 Chinese language courses for thirty minutes. The third stage involved scoring all 248 tests collected from the participants. One point was given to each correct item while 0 point was assigned to each incorrect item. Thus, the maximum score was 20 and the minimum was zero. At the fourth stage, the participants' scores were divided into three groups (beginner, intermediate and advanced) based on the cut-off points (11 between beginning and intermediate, 16 between intermediate and advanced) produced from the Angoff procedure. Accordingly, 91 students who received a score between 1 and 10 were identified as being at the beginning level; 110 students with scores between 11 and 15 were deemed as being at the intermediate level; and 47 students with scores between 16 and 20 were deemed as being at the advanced level.

ANALYSIS

Descriptive statistics of the test scores (mean, standard deviation, minimum and maximum scores, and skewness and kurtosis) were calculated to obtain information about the central tendency, dispersion, and shape of the distribution of the 248 examinees' test scores. Descriptive statistics were also calculated at the group level (beginner, intermediate, and advanced) to compare the reading performance at the three different levels. After screening the overall picture of the test scores, Confirmatory Factor Analyses (CFAs) were performed using EQS version 6.1 (Bentler & Wu, 2005) to examine the adequacy of the theoretical model of reading ability and to determine the underlying structure of test performance, as measured by the 20 reading items. CFAs are often used in validation studies because they evaluate an overall model, as well as individual parameters specified in the model. In order to evaluate the fit of the model, fit indices were examined, such as the Chi-square statistic, the Chi-square/df ratio, the comparative fit index (CFI), and the root mean-square error of approximation (RMSEA). Then, the statistical significance of the parameter estimates was checked for evaluation regarding the fit of the individual parameters.

After assessing the structure of the reading test, an Analysis of Variance (ANOVA) was used to examine whether the three groups of examinees (beginner, intermediate, and advanced) exhibited differences in each reading test component (i.e., literal, intended, and implied meaning). After confirming group differences in the reading performance for each reading test component, a stepwise regression analysis was performed to explain the proportion of variance (overall reading test scores) that could be predicted by the three test components. The R-square statistics were calculated to explain the test components that represented the test performance of each group and their contributions (Song, 2005).

RESULTS

DESCRIPTIVE STATISTICS

Descriptive statistics of the test scores were computed for each group of examinees (beginner, intermediate, and advanced) and for the entire group of examinees. The results of the descriptive statistics are presented in Table 2.

TABLE 2. Descriptive statistics for each group and entire examinees

Examinee group	# of examinees	Minimum	Maximum	Mean	SD	Skewness	Kurtosis
Beginner	91	1	10	6.74	2.30	-0.41	-0.69
Intermediate	110	11	15	13.17	1.30	-0.17	-1.07
Advanced	47	16	19	16.64	0.87	1.21	0.57
Total	248	1	19	11.47	4.18	-0.47	-0.74

The preliminary test results indicate that overall, the test had an appropriate difficulty level for these learners, resulting in a normal distribution of scores. The distributions of scores were approximately normal for each examinee group and for the entire group of examinees. However, contrary to the beginner, intermediate, and entire group, the advanced group reported positive skewness (1.21) and kurtosis (0.57) values, indicating that more examinees in the advanced group received lower scores than the mean of the group (16.64). Among the three groups, the advanced examinees showed the least variability in their scores, while the beginner examinees had the largest variability. The internal consistency reliability for the twenty item test was also estimated using Cronbach's alpha. The coefficient alpha was 0.83, which suggests that the twenty items were measuring the same construct (reading ability) to a moderate degree.

RESULTS OF CONFIRMATORY FACTOR ANALYSIS

In order to examine whether the hypothesized constructs of reading ability (literal, intended, and implied) would function as intended, a series of CFAs were conducted, and the final underlying structure of test performance was obtained. While testing the possible assumptions of the structure, items 2 and 3 were deleted from the analyses because the loadings of these two items were not statistically significant (test statistic $< \pm 1.96$). Items 2 and 3 were supposed to load on the literal and intended meaning factors, respectively, for substantive reasons; however, they did not meaningfully contribute to the model, and were thus, deleted from the model. The remaining 18 reading items loaded on the three trait factors, and the test structure followed a representation of the reading ability model. As seen in Figure 1, six items (1, 5, 7, 11, 15, and 18) loaded on literal meaning; eight items (4, 6, 8, 9, 12, 13, 16, and 19) loaded on intended meaning; and four items (10, 14, 17, and 20) loaded on the implied meaning factor. The items loaded on each factor as expected. The three trait factors (literal, intended, and implied meaning) were correlated with one another to a high degree (0.86 to 0.89).

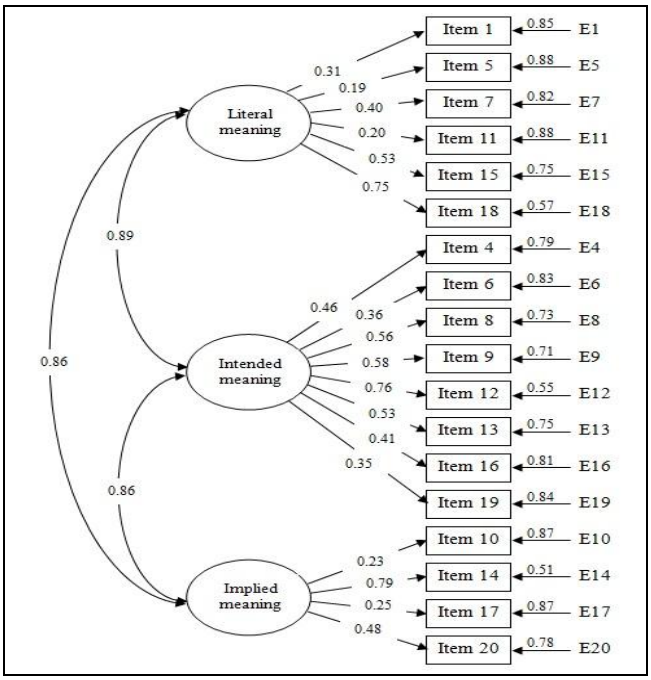

FIGURE 1. Three-factor model of the reading test

To assess the model as a whole, goodness-of-fit statistics were calculated. The independence model Chi-square statistic was 1,201.58 with 153 degrees of freedom ($p < 0.0001$), suggesting that the data did not fit the hypothesized model. However, Chi-square statistics are known to be sensitive to sample size (Byrne, 2006); thus, other fit indices were further examined. The CFI was reported as 0.903, and the RMSEA was 0.058, with a confidence interval of 0.046 and 0.069. The CFI was greater than 0.09, which indicated a well-fitting model. RMSEA values less than 0.05 are normally considered a good fit; however, values as high as 0.08 are acceptable as reasonable errors of approximation (Browne & Cudeck, 1993; Byrne, 2006). Therefore, overall, the reading performance data fit the model adequately as a whole, but not as well as anticipated.

After assessing the model fit, individual parameter estimates were also assessed. The unstandardized estimates showed that all parameter estimates (factor loadings, covariances, and error variances) were statistically significant, suggesting that they contributed to the model as important elements. Standardized parameter estimates revealed that factor loadings ranged from 0.19 to 0.79, while error loadings were estimated at approximately 0.75. The results suggest that the contribution of errors to the variables was not negligible, and that the variables could not mainly be explained only by the factors. In other words, variables other than the three factors (e.g., test method) should be considered along with the three factors (albeit unexpected) when explaining the examinees' test performance on the 18 reading items.

RESULTS OF GROUP COMPARISONS

After examining the test structure of the 248 examinees' reading performance, the differences among the three groups' (beginner, intermediate, and advanced) test scores were further analyzed. In order to compare the group differences for each of the three types of reading items (literal, intended, and implied meaning items), ANOVAs were used. First, the assumptions of ANOVA were tested, including the independence of observations, homogeneity of variances, and normality.

The results indicated that the mean scores of the three groups were significantly

different for each item type: $F(2, 245) = 198.52$ ($p < 0.001$) for literal meaning; $F(2, 245) = 386.39$ ($p < 0.001$) for intended meaning; and $F(2, 245) = 182.55$ ($p < 0.001$) for implied meaning. Since each of the three ANOVA results indicated a significant difference among the three groups, a post hoc test (Tukey's HSD test) was performed to further analyze which examinee groups differed in their reading performance. The results indicated that all three examinee groups showed a significant difference from one another for each of the three item types. Therefore, it is evident that the three groups were not equivalent with respect to their ability to comprehend literal, intended, and implied meaning. The three groups' test performances are presented for each reading item types in the following graphs.

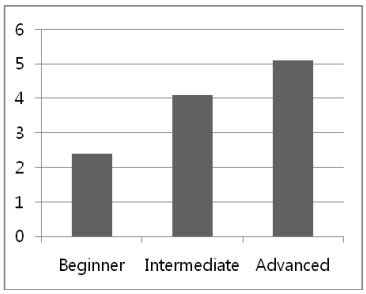

FIGURE 2.
Mean score difference
for literal meaning

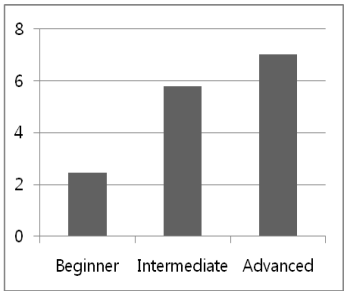

FIGURE 3.
Mean score difference
for intended meaning

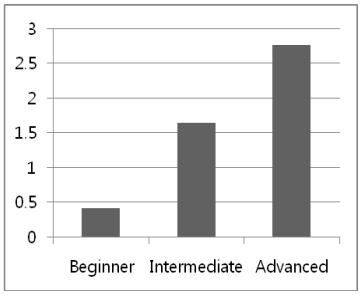

FIGURE 4.
Mean score difference
for implied meaning

In order to further explain the group differences, a stepwise regression analysis was performed for each examinee group. Tables 3, 4, and 5 present the results of the regression analysis. Only the predictors that had a significant effect on the dependent variable (total test scores) were included in the model summary. As a result, unequal numbers of models are presented for different examinee groups.

TABLE 3. Model summary for beginner examinees' reading test scores

Model	R	R-Square	Adjusted R Square
1	0.86[a]	0.74	0.74
2	0.96[b]	0.93	0.92

a. Predictors: (Constant), Literal
b. Predictors: (Constant), Literal, Intended

Beginner examinees' test scores were mainly explained by the literal and intended meaning test items. The literal meaning items alone explained 74% of the total score variance, while the literal and intended meaning items explained 93% of the variance. That is, most of the score variance was explained by the first two predictors, whereas the implied meaning items contributed almost nothing to predicting beginner examinees' test performance. Since the effects of the implied meaning items were not significant, implied meaning was not included as a predictor in the model.

TABLE 4. Model summary for intermediate examinees' reading test scores

Model	R	R-Square	Adjusted R Square
1	0.54[a]	0.29	0.28
2	0.82[b]	0.68	0.67
3	1.00[c]	1.00	1.00

a. Predictors: (Constant), Literal
b. Predictors: (Constant), Literal, Intended
c. Predictors: (Constant), Literal, Intended, Implied

Contrary to the beginner examinees' model, all three test components had to be included as predictors to explain the intermediate examinees' test score variance.

TABLE 5. Model summary for advanced examinees' reading test scores

Model	R	R-Square	Adjusted R Square
1	0.51[a]	0.26	0.25
2	0.70[b]	0.49	0.46
3	1.00[c]	1.00	1.00

a. Predictors: (Constant), Literal
b. Predictors: (Constant), Literal, Intended
c. Predictors: (Constant), Literal, Intended, Implied

Similar to the intermediate examinees' model, all three factors together explained the total score variance of the advanced examinees. However, the literal and intended meaning items could explain only about half of the score variance (49%). Therefore, the role of the implied meaning items was more important in the advanced group in predicting the advanced examinees' test performance, compared to the intermediate group.

DISCUSSION

STRUCTURE OF THE CHINESE READING TEST

The results of the CFA show that the reading test was structured according to the three types of reading (literal, intended, and implied meaning comprehension). This underlying test structure represents Purpura's (2004) meaning-based model of language knowledge. Other studies involving L2 reading ability derived from Purpura's model have usually had two underlying factors. For example, Liao (2008) operationalized L2 reading ability variables in terms of literal meaning (understanding explicitly stated information) and pragmatic meaning (understanding implicit information). Similarly, Kim A. Y. (2011) explained reading ability with semantic (literal and intended) and pragmatic (implied) meaning. Contrary to these previous studies, the test constructs of the present study were defined in terms of the three separate types of meaning (literal, intended, and implied) that Purpura (2004) originally differentiated. Also, the test items were developed to measure examinee ability so as to understand the three different types of meaning. Therefore, the CFA results provided empirical evidence of construct validity in the present study, as well as in Purpura's (2004) theoretical model of language knowledge.

While the overall fit indices and significance of all parameters confirmed that the test items measured the underlying constructs of reading ability as intended, a few compromises had to be made during the model-building process. First of all, the two insignificant items (Items 2 and 3) were deleted from the analysis. Test item number 2 (literal meaning item) provided a short paragraph in which test-takers had to choose among four options the one concurring with the content of the text. However, the phrase "拒绝别人" (*refuse other people*), which was one of the options, also appeared in the text; thus, participants apparently tended to choose this option because of its similarity with the text, not because they understood the literal meaning of the phrase. Test item number 3 (intended meaning item) required test-takers to read three sentences and then put them in the correct order. However, this sequencing could be accomplished only by understanding the first word of each sentence, rather than by truly assessing meaning between the lines. Since the two items failed to measure participants' true understanding of the literal and intended meaning as intended, the loadings of these two items might not have been statistically significant.

The other problem of the model-building process involved the relatively large contribution of error loadings, as discussed earlier. Because of these error contributions, it was difficult to conclude that the test performance was mainly explained by the underlying factors of reading ability. That is, something else was involved in the test performance other than the ability to understand literal, intended, and implied meaning. Possible reasons for this problem may be found in the test development and administration procedures. For practical reasons, only 20 items were developed, and two of these 20 were deleted during the analysis. Therefore, the number of test items was not sufficient to adequately measure the three trait factors. For example, only four items were intended to measure the participants' ability to understand implied meaning. Moreover, the sample sizes for the three different examinee groups (beginner, intermediate, and advanced) were not balanced. While the beginner and intermediate groups had 91 and 110 examinees, respectively, the advanced group had only 47 examinees. Due to these two main reasons, a few compromises were inevitable in structuring the test performance. As a result, the underlying trait structure could only partially explain participants' test performance.

GROUP COMPARISONS OF TEST PERFORMANCE

In each of the reading item types, the three groups of examinees (beginner, intermediate, and advanced) showed significant differences in their performance. At the beginning of the study, cut-scores were determined based on the Angoff (1971) method. These cut-scores were then used to create three groups from the test scores (beginner, intermediate, and advanced). We then examined the three groups' performance on the sets of items that assessed their ability to understand literal, intended, and implied meaning. According to the ANOVA results, the reading performances of all three groups were not the same across the three types of meaning items.

Further regression analyses explained how the three groups differed in their reading ability. The beginner examinees seemed to mainly understand literal and intended meaning. Most of their reading performance (93% of the variance) was based on these two types. With only the literal meaning predictor, 74% of the score variance was explained. Therefore, it can be concluded that beginner examinees could comprehend explicitly stated literal meaning from the text, while they had difficulty inferring meaning from the text or outside of the text. However, this finding is not surprising because there were no items created specifically for the beginning level that assessed implied meaning. Participants in the beginner level might have guessed answers for the implied meaning items, which were too difficult for them, and thus, the variance became random. Assessing implied meaning might only be possible at higher levels, but the absence of implied meaning items targeting the beginning level makes it difficult to conclude that beginner-level learners entirely lack the ability to understand implied meaning.

The intermediate examinees appeared to handle all three types of meaning items, as the effects of all three predictors were statistically significant and were included in the model. For the intermediate level, two literal, three intended, and two implied meaning items were developed, and each type explained 29%, 39%, and 31% of the score variance, respectively. Since all three types of meanings were treated as equally important on the test, the intermediate examinee performance results correspond to the predicted amount of score variance. Thus, this group of examinees could understand meaning that was explicitly and implicitly stated in the text and could make inferences about text meaning using their background knowledge at their reading ability level. The last group, advanced examinees, appeared to have the ability to understand any type of reading item. Their reading performance was minimally or partly explained only by the literal meaning predictor (26%,

see model 1 in Table 5) or by the literal and intended meaning predictors (49%, see model 2 in Table 5). That is, the advanced examinees' reading performance could not be fully explained without any of the three types of items. As seen in the regression analyses, the three examinee groups showed distinct performance patterns with respect to the three different types of meaning items.

In foreign language classes in Korea, reading is often taught by two approaches: extensive reading, focusing on the top-down model, and intensive reading, focusing on the bottom-up model (Chen, 2010; Zhao, 2004; Zhou & Li, 2009). Chinese reading at the beginning level most commonly involves intensive reading (Kim & Park, 2013). At this level, reading comprehension focuses on sounds, letters, words, and grammatical rules (Zhou & Li, 2009). Therefore, the general focus is on understanding the surface-level meaning attached to the text. As the present research reports, while beginning-level participants were likely to get high scores on items asking for literal meanings, they were less likely to get positive scores on items asking for implied meanings. Intermediate-level Chinese teaching and learning mostly emphasize intensive reading with some extensive reading (Kim & Park, 2013). Teaching focuses on having students establish their own reading strategies, including skimming and scanning, searching keywords and finding the gists of texts. Students are guided to understand not only the literal meaning, but also the interpretive level meaning attached to the text. Therefore, at the intermediate level, participants were found to score higher on items asking for intended meanings. At the most advanced level of Chinese teaching and learning, most reading concentrates on extensive reading involving a wide variety of texts (Kim & Park, 2013). Learners are encouraged to go beyond understanding information given in the text, and to seek the implied-level meaning attached to the text, aiming to understand the author's intentions. Therefore, advanced-level learners may have been able to enhance their skills in comprehending not only literal and intended meanings, but also implied meaning. The advanced-level learners obtained positive results in all sections, including items related to intended and implied meanings.

CONCLUSION AND IMPLICATIONS

The current study examined the underlying trait structure of the reading items developed based on the Purpura's (2004) language ability model. Overall, the test performance structure represented the theoretical model, and the items appeared to measure literal, intended, and implied meaning, as intended. A further analysis of group comparisons revealed different performance patterns across the beginner, intermediate and advanced examinee groups. Lower-level examinees' reading performance largely depended on literal meaning items, while higher-level examinees exhibited a more balanced understanding of literal, intended, and implied meaning from the text. As argued above, however, this finding might be due to the test item types created for each level.

The present research has contributed to enhancing the validity of Chinese reading proficiency assessment tools, with the aim of helping them more accurately assess learners' Chinese reading ability by presenting a theoretical foundation for developing testing items. Zhao (2004) points out that teaching and learning Chinese reading is still quite slanted toward acquiring linguistic knowledge, including vocabulary and grammar. Such teaching and learning methods may be conducive to cultivating an understanding of literal meaning, but may minimally help learners foster their comprehension of intended meanings that require reading between the lines. It is even more difficult to expect such methods to help learners understand implied meanings that require reading beyond the lines. In addition, most Chinese reading assessment has not attempted to measure learners' understanding of different types of meaning, including the New HSK, which is currently the most representative Chinese

proficiency test available. A test is used not only as a tool to determine the efficiency of language teaching and learning, but also as a guide and stimulant for finding future directions for further teaching and learning practices (washback effects) (Sadeghi & Nikou, 2012; Salehi & Yunus, 2012; Zhao, 2004). The present research suggests a direction for future teaching and learning in Chinese reading so that Chinese language learners can equip themselves with balanced skills in reading comprehension, and eventually will be able to actively exchange and share information with others through Chinese language.

LIMITATIONS AND SUGGESTIONS FOR FURTHER RESEARCH

As briefly mentioned earlier, the present study has a number of limitations. Due to the small sample size and unbalanced number of examinees for the three examinee groups, a more sophisticated statistical analysis (e.g., multi-group analysis and CFA for each examinee group) was not performed. The lack of reading items may also have made it difficult to explain the main effects of each item developed to elicit examinee ability to comprehend the different types of meaning. Therefore, a larger sample size and a more careful selection of examinees representing a wide range of reading ability may better explain Chinese learners' strengths and weaknesses in their reading, and may further inform how such strengths and weaknesses are different across ability levels. Beyond the statistical analyses, additional qualitative analyses of examinee performance, such as think-aloud protocols and interviews, would also contribute to a better understanding of learners' reading ability.

Not only in China, but also in countries where Chinese is taught as a second or foreign language, including Korea, research on Chinese testing and evaluation is quite small in number, compared to research on Chinese language teaching and learning. This is even more so with research on Chinese reading ability tests. So far, Chinese proficiency tests have been functioning as a tool to measure the efficiency of Chinese teaching and learning. As tests have been advancing to the forefront of Chinese language instruction as a guide and impetus for Chinese teaching, it is strongly anticipated that research on Chinese reading ability tests will gain momentum and will become increasingly important in the near future.

ACKNOWLEDGEMENT

This work was supported by Hankuk University of Foreign Studies Research Fund.

REFERENCES

Alderson, J. C. (1990). Testing reading comprehension skills (Part one). *Reading in a Foreign Language. 6*, 425-438.

Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.). *Educational measurement (2nd ed.)* (pp. 508-600). Washington, DC: American Council on Education.

Bentler, P. M. & Wu, E. (2005). *EQS 6.1 for windows user's guide*. Encino: Multivariate Software, Inc.

Browne, M. W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long. (Eds.). *Testing structural equation models* (pp. 136-162). Newbury Park: Sage.

Byrne, B. M. (2006). *Structural equation modeling with EQS*. Mahwah: Lawrence Erlbaum Associates, Inc.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*.

Cambridge: Cambridge University Press.

Chen, Cheng Lian (2010). *Teaching Chinese as a foreign language: Introduction*. Shanghai: Fudan University Press.

Grabe, W. (1991). Current developments in second language reading research. *TESOL Quarterly. 25*, 375-406.

Jeong, Youn Chang. (2008). Investigating the relationship between the theoretical framework of the New HSK and teaching Chinese as a foreign language. *Chinese Studies. 36*, 49-73.

Kim, Ah Young. (2011). Examining second language reading components in relation to reading test performance for diagnostic purposes: A fusion model approach. Unpublished Ed.D thesis, Teachers College, Columbia University, New York, USA.

Kim, M. S. (2011). A proposal on evaluation and teaching-learning method of communicative skills in Chinese language–Through the analysis of assessment from New HSK and SAT II. *Journal of Chinese Linguistics in Korea. 35*, 443-470.

Kim, M. & Park, Y. (2013). A study on purpose of Chinese reading proficiency test. *Journal of Chinese Language and Literature. 78*, 147-166.

Liao, Yen-Fen. (2008). A construct validation study of the GEPT reading and listening sections: Re-examining the models of L2 reading and listening abilities and their relations to lexico-grammatical knowledge. Unpublished Ed.D thesis, Teachers College, Columbia University, New York, USA.

Lumley, T. (1993). The notion of subskills in reading comprehension test: An EAP example. *Language Testing. 10*, 211-234.

Purpura, J. E. (2004). *Assessing grammar*. Cambridge: Cambridge University Press.

Sadeghi, K. & Nikou, S. B. (2012). Perceptions of Iranian high school EFL teachers and students toward teaching and learning reading skill. *3L: Language Linguistics Literature®, Southeast Asian Journal of English Language Studies. 18*(4), 167-180.

Salehi, Hadi & Melor Md Yunus. (2012). The washback effect of the Iranian universities entrance exam: Teachers' insights. *GEMA Online® Journal of Language Studies. 12*(2), 609-628.

Seong, Yoon Suk. (2010). An analysis of the new version of HSK as the tool for valuation in Chinese: Focusing on the comparative analysis between the new 5th grade HSK test and the old beginner, intermediate grade HSK tests. *The Journal of Linguistics Science. 54*, 41-82.

Song, Xioamei. (2005). Language learner strategy use and English proficiency on the Michigan English language assessment battery. *Spaan Fellow Working Papers in Second or Foreign Language Assessment. 3*, 1-26.

The Office of Chinese Language Council International. (2007). *Chinese language proficiency scales for speakers of other language*. Beijing: Beijing Foreign Language Teaching and Research Press.

The Office of Chinese Language Council International & Confucius Institute Headquarters. (2009a). *Chinese language proficiency test syllabus: Level 1*. Beijing: The Commercial Press.

The Office of Chinese Language Council International & Confucius Institute Headquarters. (2009b). *Chinese language proficiency test syllabus: Level 2*. Beijing: The Commercial Press.

The Office of Chinese Language Council International & Confucius Institute Headquarters. (2009c). *Chinese language proficiency test syllabus: Level 3*. Beijing: The Commercial Press.

The Office of Chinese Language Council International & Confucius Institute Headquarters.

(2009d). *Chinese language proficiency test syllabus: Level 4*. Beijing: The Commercial Press.

The Office of Chinese Language Council International & Confucius Institute Headquarters. (2009e). *Chinese language proficiency test syllabus: Level 5*. Beijing: The Commercial Press.

The Office of Chinese Language Council International & Confucius Institute Headquarters. (2009f). *Chinese language proficiency test syllabus: Level 6*. Beijing: The Commercial Press.

Weir, C. J. (1997). The testing of reading in a second language. In C. Clapham & D. Corson (Eds.). *The encyclopedia of language education: Vol. 7. Language testing and assessment* (pp. 39-49). Dordrecht, The Netherlands: Kluwer.

Zhao, Min. (2004). *Teaching Chinese as a foreign language: Introduction*. Beijing: The Commercial Press.

Zhou, Jing. & Li, Hoi Yeung. (2009). *The Guidance of Teaching Chinese to speakers of other languages*. Guangzhou: Sun Yat-sen University Press.

APPENDIX

试卷

第1~2题：请选出正确答案。

1. 小时候弟弟比我矮，现在却超过我了，看着他一米八二的个子，我真是羡慕极了。

★根据这句话，可以知道现在：

A 我一米八　　B 我比弟弟矮　　C 弟弟个子矮　　D 我同情弟弟

2. 有的时候，我们要学会拒绝别人。拒绝别人，要找到合适、礼貌的方法，否则，如果表达不合适，就会引起误会。

★这段话主要说怎样：

A 拒绝别人　　B 获得尊重　　C 减少误会　　D 获得原谅

第3~4题：排列顺序。

3. ①她就给我留下了极深的印象。
②那就是她特别热情、特别友好。
③第一次和王小姐见面。

A ①→②→③　　B ②→③→①　　C ③→①→②　　D ①→③→②

4. ①这种鱼生活在深海中。
②看起来像一个个会游泳的小电灯。
③它们的身体能发出美丽的亮光。

A ①→②→③　　B ②→③→①　　C ③→①→②　　D ①→③→②

第5~6题： 请选出正确答案 。

什么是真正的朋友？不同的人会有不同的理解，而我的理解是：在你遇到困难的时候，朋友会勇敢地站出来，及时给你帮助；在你孤单或者伤心流泪的时候，朋友会陪在你身边，想办法让你改到幸福；无论你是穷人还是富人，真正的朋友永远值得你信任。

5. 根据这段话，朋友可以帮我：
A 总结经验　　B 照顾家人　　C 远离危险　　D 解决难题
6. 这段话主要介绍的是：
A 精神　　B 爱情　　C 态度　　D 友谊

ISSN: 1675-8021

第7~8题：请选出与试题内容一致的一项。

7. 从1995年开始，学校每年举行一次演讲比赛，到现在已经是第15届了。今年的比赛定在下周六，对于这场比赛，我非常把握，我要争取发挥出最好水平，你们就等我的好消息吧。

A 比赛安排在周六下午　　B 我每年都参加这个比赛
C 我对这次比赛很有信心　　D 这场比赛的水平不是很高

8. 讲故事的人，往往在最吸引人的地方故意停下来。他这样做的目的，是为了引起大家的好奇心，让人有更大的兴趣听下去。讲故事的人，也可以利用这个机会观察一下大家的态度，以便接下来讲得更好。这种做法，中国人把它叫做“卖关子”。

A 卖关子可以吸引听众　　B 讲故事的人没有好奇心
C 讲故事时不应该停下来　　D 听故事的人喜欢卖关子

第9~10题：请选出正确答案 。

朋友买了一辆新车。周末，我和他一起去试车。为了试车的性能，我们把车开得很快。“我这辆车虽然不怎么有名，但速度也和那些好车差不多了吧。”朋友高兴地说。这时，前面的车突然停了，朋友急忙刹车，可是车滑行了好长的一段路才停下来，差一点儿撞到那辆车、我和朋友都吓出了一身冷汗。“现在，我终于明白一般车和好车的区别了！”朋友说。

其实，好车和一般车都可以开得很快，但它们在停车速度上却有很大的差别，好车可以更快地停下来。人生不也是这样吗？优秀的人不仅工作起来很有效率，他们也更懂得如何迅速地停下来。对于一件没有前途的事情，尽快地停下来才是最好的选择。

9. 根据上文，作者认为优秀的人和一般人的区别：

A 迅速的办事速度　　B 和别人的合作
C 办事的成果　　D 迅速的判断力

10. 作者在第一段为什么提到自己试车的经验？

A 通过举例说明自己的见解　　B 通过举例提出代表性的例子
C 通过比喻支持自己的意见　　D 通过比喻反证自己的意见

第11~14题：请选出正确答案 。

曹操得到一只大象，很想知道这只大象到底有多重。官员们都纷纷议论，发表自己的意见。有人说，制造一个巨大的秤来称。可是怎样才能造出比大象还大的秤呢？有人说，把它看成小块儿，然后再称。可是把大象杀了，知道重量又有什么意义呢？大家想了很多办法，可是都行不通。

就在这时，曹操的小儿子曹冲对父亲说：“爸爸，我有个办法可以称大象！”然后他就把办法告诉了曹操， ① ，曹操一听， ② ，立刻安排人准备称象， ③ ，并且让大家都过去观看， ④ 。

大家来到河边，看见河里停着一只大船。曹冲叫人把象牵到船上，等船身稳定时，他就在船舷与水面齐平的地方，画了一条线。然后，曹冲再叫人把象牵到岸上来。之后，他让人把

大大小小的石头，一块儿一块儿地往船上装，船身就一点儿一点儿往下沉。等船上的那条线和水面再次平齐的时候，曹冲就叫人停止装石头。官员们都睁大了眼睛，这才终于弄清了是怎么回事儿。他们连声称赞："好办法！好办法！"这时候，谁都明白，只要把船里的石头都称一下，把重量加起来，就知道这头大象有多重了。曹操___(a)___地望着众人，心里想：你们还不如我的这个小儿子聪明呢！

11. 曹冲称大象时，不使用：

A 秤　　B 江水　　C 船帆　　D 石头

12. 上文主要谈什么？

A 大象的重量　　B 曹操的选择
C 后汉的科学　　D 曹冲的智慧

13. 请选出"连连叫好"的正确位置 。

A ①　　B ②
C ③　　D ④

14. 请选出最适合的答案填空。

A 目不转睛　　B 从容不迫
C 称心如意　　D 得意洋洋

第15~17题：请选出正确答案 。

大多数的人永远都嫌自己不够有钱。然而社会学家发现，当人们真正有钱之后，又会抱怨自己没有足够的时间。从很多例子可以看出，越是有钱的人越没有时间，而穷人和那些失业的人，每天闲的难受。

人们追求财富，目的是为了让生活过得更好，可奇怪的是，人们一旦有了钱，反而更忙碌，更无法舒舒服服地过日子。

当生活不富裕的时候，很多人都想过"等我有了钱以后就可以怎么样怎么样"。在人们的想象中，"有钱"代表自由、独立、随心所欲——夏天可以到海边度假；冬天可以到山上去滑雪。

然而，当人们果真富有了，却发现自己根本无法去实现这些梦想——理由永远只有一个："没时间！"不少高收入的人，几乎都是工作狂。

看来，"有钱"和"有闲"永远难以两全。难怪有人说："当你年轻，没钱时，希望能用时间去换金钱；当你有钱后，却很难再用金钱买回时间。"

15. 人们在没钱的时候想的是什么？

A 有了钱怎么用　　B 用金钱买回时间
C 怎样得到更多的钱　　D 忙的日子是什么样的

16. 上文主要谈什么？

A 穷人和富人　　B 时间和金钱
C 理想和现实　　D 闲人和忙人

ISSN: 1675-8021

17. 根据上文，不富裕的人能想什么？

A 工作和空闲均匀分配　　B 具体地定好赚钱的目的
C 不抱怨经济上的不富裕　　D 实现梦想不得不需要牺牲

第18~21题：请选出正确答案 。

一个年轻人获得一份销售工作，勤勤恳恳干了大半年，却接连失败。而他的同事，个个都干出了成绩。他实在忍受不了这种痛苦。在总经理办公室，他惭愧地说，可能自己不适合这份工作。“安心工作吧，我会给你足够的时间，直到你成功为止。到那时，你再要走我不留你。”老总的宽容让年轻人很感动。他想，总该做出一两件像样的是之后再走。

过了一年，年轻人有走进了老总的办公室。这一次他是轻松的，他已经连续7个月在公司销售排行榜中高居榜首。原来，这份工作是那么适合他！他想知道，当初老总为什么将自己继续留用呢。

“因为，我比你更不甘心。”老总的回答出乎年轻人的预料。老总解释道：“当初招聘时，公司收到100多份应聘材料，我面试了20多人，最后却只录用了你。如果接受你的辞职，我无疑非常失败。我深信，既然你能在应聘时得到我的认可，也一定有能力在工作中得到客户的认可，你缺少的只是机会和时间。与其说我对你仍有信心，不如说我对自己仍有信心。”

我就是那个年轻人。

18. 老总当初为什么要留这个年轻人？

A 公司急需人员　　B 客户欣赏年轻人
C 相信自己没有看错人　　D 年轻人有丰富的工作经验

19. 根据上文，不一致的内容是：

A 应征的竞争率非常高　　B 应聘的过程需要一个步骤
C 每个月公布职员的业绩　　D 客户的评价和职员的业绩有关联

20. 上文主要想告诉我们：

A 好领导能决定公司成败　　B 成功离不开集体的支持
C 工作中要学会为人处世　　D 自信和宽容成就新天地

English Translation

1-2. Choose the correct answer.

1. My brother used to be shorter than me when we were young, but now he has grown taller. His height is 182 cm, which makes me feel jealous.

★Based on the statement above, the current situation is that:

A) I am 180 cm tall.　　B) I am shorter than my brother.
C) My brother’s height is short.　　D) I feel sympathy for him.

2. We must learn how to make rejections. When rejecting others, it is necessary to find the

appropriate and polite way. When the expression is improper, it is easy to cause misunderstanding.

★The best answer that summarizes this paragraph is:

A) How to reject others B) How to get respect
C) How to prevent misunderstanding D) How to ask for forgiveness

3-4. List the answers in the right order.

3. (1) She has left me a strong impression.
(2) It is because she is very passionate and polite.
(3) I met Miss Wang for the first time.

A) (1)→(2)→(3) B) (2)→(3)→(1)
C) (3)→(1)→(2) D) (1)→(3)→(2)

4. (1) This type of fish live deep in the sea.
(2) They look like a group of swimming light bulbs.
(3) They have a luminous body.

A) (1)→(2)→(3) B) (2)→(3)→(1)
C) (3)→(1)→(2) D) (1)→(3)→(2)

5-6. Choose the correct answer.
What does 'a true friend' mean? Everyone has a different thought. In my opinion, a true friend is courageous enough to help you when you face a difficulty; a true friend stays with you to make you happy when you feel lonely or devastated; a true friend is always dependable whether you are rich or poor.

5. According to the passage, 'a true friend':
A) Shares experience B) Cares for your family
C) Pulls you away from danger D) Leads you through hard times

6. The main topic of this passage is
A) Spirit B) Love C) Attitude D) Friendship

7-8. Choose the best answer that corresponds with the given passage.

7. My school has been holding an annual speech contest since 1995. This year is the 15th time the contest is held. This year's contest is on Saturday and I am certain that I will receive a good result. I will outperform with a high level of skillset. Therefore, you should wait to hear my good news.

A) The contest is scheduled on Saturday. B) I participate in this contest every year.
C) I am confident about this contest. D) The level of this contest is not too high.

8. Sometimes, a storyteller purposely goes silent for a moment when people are absorbed in the story. The purpose of this behavior is to arouse curiosity from the audience and make

them listen more carefully. In order to create a more pleasant atmosphere, the storyteller uses the short breaks to observe the listeners' reaction. Chinese people call this way of speaking, 'intended pausing.'

A) Intended pausing attracts the audience's notice.
B) The storyteller has no curiosity.
C) When telling a story, one should not go silent at any time.
D) The audience is favorable of intended pausing.

9-10. Choose the correct answer.

My friend purchased a new car. Last weekend, he and I went out for a test drive. In order to test the car's performance, we drove fast. My friend said in excitement, "Although this car is not so famous, its speed is as fast as luxury cars." At that moment, the car in front of us suddenly stopped, so my friend had to step on the brake abruptly. Our car could finally stop after skidding for a long time. As we could have almost crashed into another car, we were very terrified. "Now I know the difference between regular and luxury cars," said he.

In fact, both regular and luxury cars can run in high speed, but there is a big gap in stopping the car. Luxury cars usually stop more quickly than regular cars. This gap also applies to life. Smart people not only work efficiently but also know how to stop promptly. The best solution for a situation that seems to have no future is to stop as quickly as possible.

9. According to the passage, the author considers the main difference between smart people and regular people to be:

A) Quick completion of task
B) Cooperation with others
C) Achievement of tasks
D) Quick judgment

10. The author mentions his own experience in the first paragraph in order to:

A) Explain his opinion through exemplification.
B) Give a typical example through exemplification.
C) Propose his opinion through metaphor.
D) Disprove his opinion through metaphor.

11-14. Choose the correct answer.

Chao Cao wished to know how much the elephant he recently received weighed. His officials brought forward various opinions. One official suggested producing a huge scale, but it was impossible to produce such scale that was as big as the elephant. Another offered to cut the elephant into pieces and measure its weight; however, there was no meaning of measuring the elephant's weight if it died. Though many people proposed a diversity of ideas, Chao Cao was not satisfied.

Then, Chao Cao's youngest son, Chao Chong, said, "Father, I know how to measure the elephant's weight." As Chao Chong explained the method, (1) Chao Cao listened to him, (2) gave orders to prepare for the measurement, (3) gathered crowd to watch, (4).

People went by the river and there was a ship floating. Chao Chong gave orders to put the elephant on the ship, to wait until the ship stays still, and to draw a line on the side of the ship that meets horizontally with the water. Chao Chong told people to take the elephant off the ship and to fill the ship with rocks of various sizes. Then, the ship began to sink little by little. As the surface of water became aligned with the line drawn on the side of the ship, Chao Chong stopped people from filling the boat with rocks. Finally, the officials had eyes wide

with amazement to realize what had happened. They were marveled, saying "What a great method! What a great method!" After all, everyone knew that the weight of rocks that filled the ship was equal to the weight of the elephant. Chao Cao looked at the officials, (a): 'Not any one of you is as smart as my youngest son.'

11. Chao Chong made use of the following in weighing the elephant, EXCEPT
A) Scale B) Water C) Anchor D) Rock

12. The main topic of this passage is
A) Weight of the elephant B) The decision of Chao Cao
C) Science of the Last Han Dynasty D) Wisdom of Chao Chong

13. The phrase "let out exclamations" can be best placed in
A) (1) B) (2) C) (3) D) (4)

14. Choose the most appropriate phrase that replaces (a).
A) Staring B) Keeping calm C) Striving D) Flaunting

15-17. Choose the correct answer.

Most people always complain about shortage of money. Sociologists discovered the fact that when people actually have money, they tend to complain about shortage of time. As seen through several examples, the more money one has, the less time he has; the poorer one is or if one has no job, the more painful he is with boredom.

People pursue wealth in order to live a better life. However, once people gain wealth, they tend to become busier and eventually cannot live a better life.

When people confront shortage of money, many think, "If I have enough money, I would do …." In people's minds, the term "wealthy" represents freedom, independence, doing-whatever-they-want – for example, sunbathing by the shore in summer and going for mountain ski in winter.

However, when people are really wealthy, they realize that their plans cannot come true. The one and only reason is: "There is no time!" Moreover, people with high income are workaholic.

In conclusion, being wealthy and having enough time cannot be achieved at the same time. Thus, people say, "When you are young and poor, you wish to earn money with time. But, even if you have money in the future, you cannot buy time with money."

15. What do people think when they have no money?
A) Make plans for when they have enough money B) How to buy time with money
C) How to earn more money D) Imagine how a busy life is like

16. The main topic of this passage is
A) The rich and the poor B) Time and money
C) Ideal and Reality D) People with enough time and people with lack of time

17. According to the passage, people with shortage of money think that
A) work and time can be attained at the same time.
B) the purpose of earning money should be clear.
C) not having enough money is dissatisfactory.
D) making dreams come true is followed by sacrifice.

18-21. Choose the correct answer.

A young man had a job related to marketing. Although he worked hard for half a year, there was no successful achievement. However, each of his colleagues was comparatively successful. The man could not stand the pain of failure. At his boss's office, the man told him with shame that he did not fit into the job. "Do not mind other factors when you work because there is enough time until you can finally succeed. If you want to leave by then, I will let you go." The generosity of his boss touched the man's heart. The man thought that he would leave after achieving at least one successful task.

One year after, the man went into the boss's office. This time, he walked in with a light heart as he had been on top of the sales chart for seven consecutive months. In fact, the job was fit to the man. He wanted to know the reason why his boss did not fire him in the past.

"It's because I would have felt more resentful if I let you go back then." The boss's response was unexpected. The boss explained, "I received over 100 applications when the company was recruiting, interviewed about 20 applicants, and finally hired you. If I had let you go, it would mean a big failure in my career as well. I firmly believe that you could be acknowledged by the customers as I already acknowledged you and that you just did not have enough time and chance. Actually, I rather had trust in myself than in you."

This is the story of a young man, and I AM the young man.

18. The reason why the boss did not let the man go is because

A) the company urgently needed employees.
B) customers favored the young man.
C) the boss believed that his decision was right.
D) the young man had various work experiences.

19. According to the passage, the following statements are true EXCEPT

A) The competition rate was very high.
B) The recruiting session is consist of only one step.
C) The company announces each employee's achievement every month.
D) The customers' feedback is related to the employees' achievements.

20. The main topic of this passage is

A) a good leader guides the company to success.
B) success is followed by the support of majority.
C) one needs to learn how to manage relationships with people at work.
D) confidence and generosity creates a good outcome.

ABOUT THE AUTHORS

Mi Soon Kim earned a Ph.D. in Linguistics and Applied Linguistics from Beijing Language Culture University. She is an assistant professor of Chinese education in the Graduate School of Education at Hankuk University of Foreign Studies, Korea. Her research interests include teaching Chinese as a foreign language and grammar studies.

Hyun Jung Kim earned an Ed.D. in Applied Linguistics from Teachers College, Columbia University. She is an assistant professor of the Graduate School of TESOL at Hankuk University of Foreign Studies, Korea. Her research interests include language assessment and language test validation.

ISSN: 1675-8021