

An Improved Measure of Retention in Vocabulary Learning and Its Application*

Hyung-Sun Kim**

Chosun University

Ung-Yong Choi

Chonnam National University

Kim, Hyung-Sun & Choi, Ung-Yong. (2017). An improved measure of retention in vocabulary learning and its application. *Modern English Education*, 18(1), 1-24.

This study proposes a new model to measure learners' vocabulary retention, namely the Tracing Measure of Retention (TMR), as opposed to the conventional method, named as the Simple Comparison of Scores (SCS) in this study. The SCS has been mostly based on a fixed group of target words, which does not properly reflect participants' individual differences. Moreover, the delayed-test scores may possibly include the words that are not subject to retention. This study attempts to justify the newly proposed model as a solution to these issues by applying both models to an authentic learning activity and comparing the two sets of outcomes. Seventy eight university students participated in an in-class reading activity which was designed to boost incidental vocabulary learning, and the vocabulary tests accompanied. The two measurement models yielded different results, which were compared by means of participants' proficiency levels as a verification tool. The Pearson correlation analysis and ANOVA tests were carried out to check for statistical significance. The results supported the TMR as a more accurate and practical measure of retention, which would serve more insightful research.

**[tracing measure of retention (TMR)/customized target-word identification (CTI)/
기억 추적 측정/개별 목표단어 식별]**

* This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2016S1A5A8019628). Part of this research was presented at the Korean Institute of Smart Media Conference in Busan, April 29-30, 2016.

** First author: Hyung-Sun Kim, Corresponding author: Ung-Yong Choi

I. INTRODUCTION

A certain level of vocabulary knowledge is prerequisite to communication in a second language (L2); a large body of research in L2 acquisition has been devoted to this area, taking either a direct or indirect approach to vocabulary learning. Vocabulary, especially in an L2 context, was deemed more “a matter of rote-learning or memorization chore traditionally, where only an autonomous, perseverant learner would flourish” (H. S. Kim & Y. H. Na, 2010, p.184). Recently, an increasing number of studies has explored the indirect aspect, i.e., incidental development of vocabulary knowledge through various learning activities other than those focusing explicitly on vocabulary learning, such as reading (Hemmati & Asmawi, 2015; M. H. Ko, 2012) and listening (Brown, Waring, & Donkaewbua, 2008).

Reading in both L1 and L2 is broadly considered an excellent means for vocabulary acquisition. In fact, scholars agree that the quantity of reading, rather than that of spoken input and subsequent oral practice, is the major contributor to individual differences in learners’ vocabulary competence (Cunningham & Stanovich, 2001; Hayes, 1988). Others claim that learners often fail to spontaneously “pick up” the meanings of unknown words encountered in texts, and even more so in L2 acquisition (Hulstijn, Hollander, & Greidanus, 1996). Many studies have dealt with this issue, investigating the influence of various ways of pedagogical intervention to aid vocabulary learning through reading activities, such as glossing of texts (M. H. Ko, 2012), manipulation of the structure of texts (Larsen-Freeman & Long, 1991), and controlling learning task involvement load (Y. Kim, 2008; Laufer & Hulstijn, 2001).

Many such studies conducted a set of immediate and delayed tests to investigate learners’ immediate vocabulary learning and the effect of time on retention. In many cases, learners’ vocabulary retention appears to be simply measured by means of score differences between immediate and delayed tests on particular target words. In most cases, the target words have been those that participants declared to be “unknown,” or those identified so by a pilot test completed by a group with similar characteristics to the target participants, who were themselves not participants in the study. Despite significant findings that may have led to valuable educational insights, the measurement may leave room for doubt as to the reliability of findings, as (a) participants may be mistaken about their previous knowledge on words; and (b) delayed-test scores may reflect correct answers to words that were answered incorrectly in the immediate test due to chance or other reasons, and thus should not be regarded as reflecting retention.

The present study aims to address these issues, as accurate measurement of retention is one of the keys to designing L2 activities that accelerate incidental vocabulary development of learners. It intends to propose a new methodology that may improve

precision and accuracy in measuring L2 learners' immediate acquisition and retention of vocabulary based on the results of this study. For an empirical comparison of the new and conventional measurement in terms of adequacy, an in-class reading activity was chosen for two reasons: (a) reading is a widely recommended means for incidental vocabulary learning (Brown, Waring, & Donkaewbua, 2008); and (b) because participants are under the guidance and supervision of the instructor, their learning time and attention can be controlled.

An illustrative example is provided to clarify the distinction between the prevailing and new methods, and to make the new method easily accessible for researchers who might want to try it. The significance of the differences between the two methods will be demonstrated through the statistical analysis of participants' vocabulary retention. In sum, the objective of this study is two-fold: (a) to present a new model to measure vocabulary gain (retention); and (b) to demonstrate the new model's greater accuracy by comparing two sets of outcomes of applying both the new and traditional models to actual data.

II. RESEARCH BACKGROUND

1. Previous Studies on Vocabulary Learning

Two approaches to L2 vocabulary development may be identified from previous research: incidental development (Coady, 1997; Hulstijn, Hollander, & Greidanus, 1996; S. Lee, 2013; Yoshii & Flaitz, 2002), i.e., an enhanced vocabulary as a "byproduct" of some activities serving other needs; and intentional (explicit) learning (Hunt & Beglar, 2002), i.e., taking relevant classes and/or memorizing words. Scholars seem to agree that after learners pass a certain "threshold" vocabulary level, i.e., the first few thousand words that will serve in most contexts due to their high frequencies (H. S. Kim & Y. H. Na, 2010), incidental vocabulary learning becomes dominant, as they guess at the meaning of unknown words encountered while reading (Huckin & Coady, 1999), for example.

Thus, a mainstream research interest in vocabulary acquisition seems to focus on ways in which input modification may enhance incidental vocabulary learning. Input modification is "a pedagogical intervention in which teachers manipulate" learning materials to help learners acquire target forms (M. H. Ko, 2012, p. 57). In research aiming to verify the positive effects of a type of input modification, accurate measurement would be of crucial importance. Several such studies are reviewed with a focus on the process of target word identification and the administration of immediate and delayed tests in order to illustrate the problems about measuring learner's vocabulary retention raised in Section I; whichever type of input modification previous studies experimented, most of them

employed the conventional method that are not free from those issues.

One input modification technique shown by Hulstijn, Hollander, and Greidanus (1996) to improve vocabulary learning is marginal glossing. These researchers reported that frequency of occurrence fostered learning when marginal glosses were provided among 78 Dutch university students learning French. Sixteen words “estimated to be unfamiliar to most first-year students” (p. 329) were targeted in the study. Despite the prudent research design in terms of scoring participants’ performance and checking their recognition and pre-knowledge, the potential inclusion of already-known words as target cannot be discounted.

Yoshii and Flaitz (2002) examined the effects of three kinds of annotation as a form of input modification in multimedia reading using the Internet: text-only, picture-only, and a combination of the two. They reported that, among 151 adult learners of English as a foreign language (EFL) at beginning and intermediate levels, the group with the combination of text and picture annotations outperformed the other two in both pre- and posttests. Fourteen target words were used in the study, and any participant who declared previous knowledge of more than 30% of the fourteen targets in the pretest were excluded. However, the inclusion of those familiar with less than 30% of the target words may have had a moderating effect on the results of subsequent immediate and delayed posttests.

M. H. Ko (2012) compared L1 and L2 glosses as input modification, along with no-gloss condition, in L2 vocabulary learning among 90 Korean university students, and reported that a significant difference was found between glossed and no-glossed texts, but not between L1 and L2 glossed texts. Her methodology was creative, in that she employed weighted scoring for the immediate and delayed tests to avoid the potential triggering of higher scores by matches between learning and testing methods. However, the possibility of already-known words being included as target words was not eliminated. Sixteen target words were selected following a pilot study with 25 other students, namely those “words unknown to 60% or more of the students” (p. 35). Thus, words known to 40% or less of the students were included.

Whereas the majority of studies on vocabulary learning have paid attention to the type of learning tasks that elicit incidental development, most notably reading, others have raised a different issue regarding tasks, namely that of involvement load (IL) (Hulstijn & Laufer, 2001). H. S. Kim and Y. H. Na (2010) corroborated the IL hypothesis in their research with 97 Korean university students learning EFL. They reported that the writing group (the high IL group), outperformed the gap-fill group (the medium IL group), which in turn outperformed the reading group (the low IL group). They used ten target words that appeared to be difficult for average university-level EFL learners, yet there remains a chance that some participants may have known (at least partially) some of the words.

Another stream of research on vocabulary learning has suggested that “growth of word

knowledge is incremental,” probably over multiple episodes (Frishkoff, Perfetti, & Collins-Thompson, 2011, p.72). In this vein, Hemmati and Asmawi (2015) compared the partial and full word-knowledge gained by 30 Iranian EFL students following reading tasks. These researchers measured gains by means of word-form/meaning recognition and meaning translation and reported significant gains at all levels but to different extents. The experiment on different levels of word knowledge was creative, yet the measurement in immediate and delayed tests was made on the basis of 25 target words of which the participants’ previous knowledge was not revealed.

Regardless of the types of learning activities investigated, the cases reviewed above suggest that many studies of vocabulary learning rely on either the participants’ declaration of their previous vocabulary knowledge or an assumption of homogeneity between experimental and similar groups in identifying target words. A more fundamental issue is that the delayed tests that supposedly measure vocabulary retention do not actually track the words initially acquired through the relevant learning activities. Most studies repeated the same measure for immediate and delayed tests. In the following section II.2, we propose a new research methodology that could resolve these issues, and present an illustrative example that future studies may want to replicate. The application of the proposed model vs. the conventional model is illustrated in Section III and IV on the basis of an actual case of incidental vocabulary learning through in-class reading activities.

2. A New Model to Measure Vocabulary Development: Tracing Measure of Retention (TMR)

Learners’ vocabulary competence may be associated with various aspects such as their prior exposure to words, fields of interest, and needs, etc., which are highly individual. Moreover, as their vocabulary knowledge is often partial, they might be unsure whether they knew certain words when asked the meanings of those words. Therefore, recruiting learners of similar proficiency levels as a pilot group, or asking participants about their previous knowledge to determine suitable target words for experiments as in much research, is a less than satisfactory solution. As for evaluating vocabulary retention, as reviewed in Section II.1, the prevailing practice is to compare scores on immediate and delayed posttests based on the same target words, which may raise concerns as the delayed test does not trace the retention of previously acquired words.

This section presents a new model to measure learners’ vocabulary development on the basis of customized target words. The model calculates gains by tracing from initial learning to eventual retention, and is named the tracing measure of retention (TMR). In

contrast to the prevailing conventional methodology, named the simple comparison of scores (SCS)¹ in this study, the TMR model comprises two processes, namely customized target-word identification (CTI) and the tracing measure of retention (TMR).²

The SCS typically starts with a fixed group of target words identified through a screening process, as shown in Figure 1. The immediate test (IT) and delayed test (DT) both involve the same target words, and their scores are simply compared to measure vocabulary retention. On the other hand, the TMR model is designed to allow for learners' individuality in terms of vocabulary knowledge. The procedure begins with an additional pre-knowledge test (PT) based on a relatively larger pool of potential target words. The IT and DT are both based on the same pool of target words, but their scoring is fundamentally different. Participants do not indicate previously known or unknown words. Instead, the IT tracks only the item that an individual participant answered incorrectly in the PT, ruling out those answered correctly as they reflect pre-knowledge. The scoring system is programmed to pick up the relevant items only, resulting in individually customized target words for each participant. The DT, in turn, tracks only those items answered correctly in the IT, excluding those answered incorrectly, as those not acquired at the first stage are not subject to retention. The SCS and TMR models are compared in Figure 1.

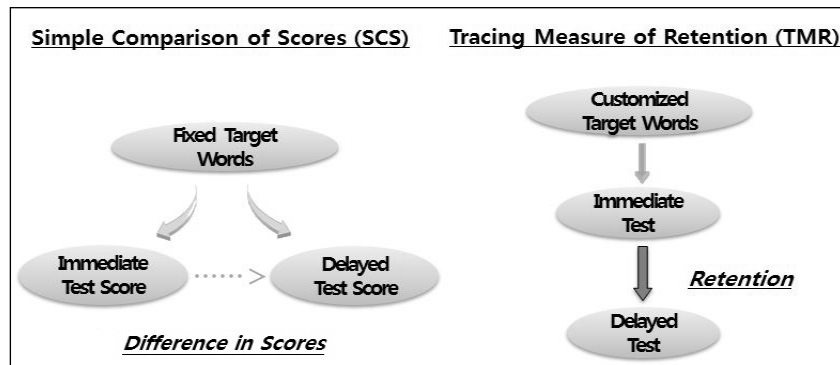


FIGURE 1 SCS vs. TMR

Table 1 represents the TMR model in terms of the imaginary performance of a hypothetical participant. W1 through W10 comprise the pool of potential target words

¹ The term simple comparison of scores (SCS) is used in two ways: (a) as the name of the conventional model; and (b) in the sections reporting the quantitative results, as the name of a variable indicating the retention ratio under the SCS model.

² The term tracing measure of retention (TMR) is used in three ways: (a) as the name of the new model proposed; (b) as the name of a process in the model; and (c) in the sections of the quantitative study, as the name of a variable indicating the retention ratio under the TMR model.

from which the customized target words (V1) for an individual participant are identified through the PT (i.e., CTI). V2 reflects performance on the IT, which is completed immediately after the relevant treatment; V4 reflects the DT, completed later to evaluate participants' vocabulary retention. In this illustrative example of TMR, the participant answered four items correctly in the PT, leaving six out of the ten words as targets (V1). Among the six targets, the participant answered four items correctly in the IT (V2), which reflects initial learning. Retention, by definition, should only include words retained from V2 to V4, namely W4 and W9 in this example. In other words, the participant initially acquired four out of the six target words, and retained two out of the four initially acquired words after a given period of time.

TABLE 1
TMR: Vocabulary Performance of a Hypothetical Participant
(0: Unknown Word, 1: Known Word)

	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10
Pre-knowledge Test (PT)	0	1	1	0	1	0	1	0	0	0
Immediate Test (IT)	1	0	1	1	0	0	1	1	1	0
Delayed Test (DT)	0	0	1	1	1	0	0	0	1	1

TMR as opposed to SCS

V1 = 6 (PT: customized target words: W1, W4, W6, W8, W9, W10)
V2 = 4 (IT: number of correct words among [V1]: W1, W4, W8, W9)
V3 = 3 (DT: number of correct words among [V1]: W4, W9, W10)^a
V4 = 2 (DT: number of correct words among [V2]: W4, W9)
V22 = 4/6 (V2/V1: score of immediate test)
V33 = 3/6 (V3/V1: score of delayed test)
SCS = 1/6 (V22 - V33)
TMR = 2/4 (V4/V2: retention)

^a The italicized part illustrates how the currently prevailing SCS works differently from the TMR.

As the SCS does not trace the initially acquired vocabulary, its DT result is based on the total target words (V1), rather than on V2. As a result, the score of the DT (V3) includes W10 that was not previously acquired, and therefore is not subject to retention. V3, which is supposedly indicative of the retention ratio, could be misleading under SCS, as Figure 2 demonstrates.

In Figure 2, the square reflects the total pool of ten target words, and the circles represent the results of the three tests, PT (upper circle), IT (lower left), and DT (lower right), respectively. The four words in the upper circle should be excluded from the actual target words as the participant already knows them. Out of the remaining six target words, the light-shaded part indicates those that are scored only in the IT, and the dark-shaded part

shows those scored in both the IT and the DT, i.e., the retained vocabulary. The Venn diagram for the SCS illustrates that W10 is incorrectly included.

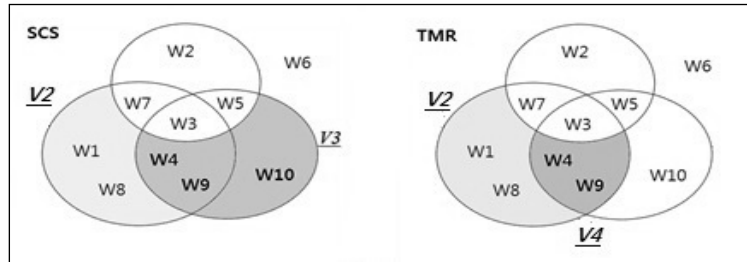


FIGURE 2 Retained Words With SCS vs. TMR

As noted, the TMR has the capacity to reflect the individuality of learners' vocabulary competence by tracing words from the PT to the IT and DT. The task is accomplished simply by utilizing Microsoft's Excel program. First, every word in the three tests is binary coded in an Excel worksheet with 1 being correct and 0 being incorrect. Second, the IF and IF AND formulas are employed to obtain the scores of the IT and DT. Figures 3 to 6 show how the formulas work to obtain the value of each variable. For example, IF(AA3 = 0, BA3) in Figure 3 indicates that, if the value of W1 is 0 in the PT, then add the value of W1 in the IT; IF(AND(AA3 = 0, BA3 = 1), CA3) in Figure 6 indicates that, if the value of W1 is 0 in PT and 1 in IT, then add the value of W1 in DT. Figure 5 shows how the DT score would be calculated by the SCS method if binary coding were followed, so that the formulas for SCS and TMR can be compared to make the difference clear.

AO3		=10-SUM(AA3:AJ3)																																
AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AO	BA	BB	BC	BD	BE	BF	BG	BH	BI	BJ	BU	CA	CB	CC	CD	CE	CF	CG	CH	CI	CJ	DE	DF	
Part 1: Pilot Scheme [0=incorrect (TARGET), 1=correct (NON-TARGET)]										Part 2: Intermediate Test [0=incorrect, 1=correct]										Part 3: Delayed Test [0=incorrect, 1=correct]														
1	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	V2	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	V2	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	V3	V6
2	0	1	1	0	1	0	1	0	0	0	6	1	0	1	1	0	0	1	1	1	0	4	0	0	1	1	1	0	0	0	1	1	3	2

FIGURE 3 Formula for the Number of Customized Target Words (V1)

BU3		=IF(AA3=0,BA3)+IF(AB3=0,BB3)+IF(AC3=0,BC3)+IF(AD3=0,BD3)+IF(AE3=0,BE3)+IF(AF3=0,BF3)+IF(AG3=0,BG3)+IF(AH3=0,BH3)+IF(AI3=0,BI3)+IF(AJ3=0,BJ3)																																
AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AO	BA	BB	BC	BD	BE	BF	BG	BH	BI	BJ	BU	CA	CB	CC	CD	CE	CF	CG	CH	CI	CJ	DE	DF	
Part 1: Pilot Scheme [0=incorrect (TARGET), 1=correct (NON-TARGET)]										Part 2: Intermediate Test [0=incorrect, 1=correct]										Part 3: Delayed Test [0=incorrect, 1=correct]														
1	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	V2	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	V2	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	V3	V6
2	0	1	1	0	1	0	1	0	0	0	6	1	0	1	1	0	0	1	1	1	0	4	0	0	1	1	1	0	0	0	1	1	3	2

FIGURE 4 Formula for the Raw Score of the Immediate Test (V2)

DE3		f _i = IF(AA3=0,CA3)+IF(AB3=0,CB3)+IF(AC3=0,CC3)+IF(AD3=0,CD3)+IF(AE3=0,CE3)+IF(AF3=0,CF3)+IF(AG3=0,CG3)+IF(AH3=0,CH3)+IF(AI3=0,CI3)+IF(AJ3=0,CJ3)	
	AA AB AC AD AE AF AG AH AI AJ AO	BA BB BC BD BE BF BG BH BI BJ BU	CA CB CC CD CE CF CG CH CI CJ DE DF
1	Part 1: Pilot Scheme [0=incorrect (TARGET), 1=correct (NON-TARGET)]		Part 2: Intermediate Test [0=incorrect, 1=correct]
2	W1 W2 W3 W4 W5 W6 W7 W8 W9 W10 V2	W1 W2 W3 W4 W5 W6 W7 W8 W9 W10 V2	W1 W2 W3 W4 W5 W6 W7 W8 W9 W10 V3 V6
3	0 1 1 0 1 0 1 0 0 0 6	1 0 1 1 0 0 1 1 1 0 4	0 0 1 1 1 0 0 0 1 1 3 2

FIGURE 5 Formula for the Score of the Delayed Test for SCS (V3)

DF3		f _i = IF(AND(AA3=0,BA3=1),CA3)+IF(AND(AB3=0,BB3=1),CB3)+IF(AND(AC3=0,BC3=1),CC3)+IF(AND(AD3=0,BD3=1),CD3)+IF(AND(AE3=0,BE3=1),CE3)+IF(AND(AF3=0,BF3=1),CF3)+IF(AND(AG3=0,BG3=1),CG3)+IF(AND(AH3=0,BH3=1),CH3)+IF(AND(AI3=0,BI3=1),CI3)+IF(AND(AJ3=0,BJ3=1),DJ3)	
	AA AB AC AD AE AF AG AH AI AJ AO	BA BB BC BD BE BF BG BH BI BJ BU	CA CB CC CD CE CF CG CH CI CJ DE DF
1	Part 1: Pilot Scheme [0=incorrect (TARGET), 1=correct (NON-TARGET)]		Part 3: Delayed Test [0=incorrect, 1=correct]
2	W1 W2 W3 W4 W5 W6 W7 W8 W9 W10 V2	W1 W2 W3 W4 W5 W6 W7 W8 W9 W10 V2	W1 W2 W3 W4 W5 W6 W7 W8 W9 W10 V3 V6
3	0 1 1 0 1 0 1 0 0 0 6	1 0 1 1 0 0 1 1 1 0 4	0 0 1 1 1 0 0 0 1 1 3 2

FIGURE 6 Formula for the Raw Score of the Delayed Test for TMR (V4)

3. Language Competence and Incidental Vocabulary Learning

In order to demonstrate the adequacy of the new model TMR, this study compares it to the conventional SCS model in actual research that aimed to investigate incidental vocabulary learning and retention following classroom reading activities. The application of the models will reveal whether the measures produce different results, and which may be better than the other.

As vocabulary retention in the research is to be measured in two ways, by SCS and TMR, and the advantage of one measure over the other is to be evaluated, verification requires a clear standard for judgement, i.e., a factor most likely to influence vocabulary retention. Therefore, learners’ scores of the College Scholastic Ability Test (CSAT) were chosen as an index of level of English competence.

The CSAT is conducted on annual basis by the Korea Institute for Curriculum and Evaluation (KICE), commissioned by the Korean Ministry of Education (KMOE) “to measure the students’ achievement of National curriculum standards and scholastic ability required for college education.”³ It covers Language Arts, Math, English, Social Studies, Science, and Vocational Education. The CSAT is mandatory for college application in Korea and is one of the exams with the greatest public confidence. The English section tests 45 items involving 17 listening and 28 reading items in multiple-choice format. The CSAT score is a standard score with the linear transformation method with a mean of 100

³ The information is available at <http://www.kice.re.kr/sub/info.do?m=0205&s=english>.

and a standard deviation of 20, resulting in stanines that are typically referred to as levels.

A number of previous studies have identified learners' language competence as an indicator of incidental vocabulary learning (Jenkins, Stein, & WYsocki, 1984; Nemati, 2013; Tekmen & Daloglu, 2006). Foreign language learners' proficiency levels were reported to be correlated to their vocabulary reception and production levels (Yuksel & Kavanoz, 2010). Indeed, it is a common understanding that high performing learners show markedly better abilities to plan their learning both consciously and unconsciously, whereas lower performers appear to struggle more when a task is assigned (H. Jeong, Biswas, Johnson, & Howard, 2010). Thus, the present methodology was aimed at taking this factor into consideration. The following research questions guided the quantitative comparison of the two models that is reported in Sections III, IV and V.

- 1) Does CSAT level significantly influence learners' vocabulary development (retention) following an in-class reading activity?
- 2) Do the SCS and TMR models yield different results for vocabulary-retention ratio?
- 3) If the answer to the question 2 is yes, which of the models is more plausible?

III. METHOD

1. Participants

The participants were 78 students enrolled in a first-year obligatory College English course at a four-year university in Korea, offered at two different levels for the general line and the art, music, and physical education line. They had various majors, including Economics and Business Administration (17), Korean and Korean Education (12), Math Education (6), Physics Education (10), Applied Chemistry and Chemistry Education (10), Electronic Engineering (8), and Premed (15), so all belonged to the general line. From the information gathered through a questionnaire,⁴ Table 2 shows the participants' CSAT and self-rated proficiency levels. At the university where the current research was carried out, Levels 1 and 2 out of the nine levels are generally considered advanced, Levels 3 and 4 as intermediate, and Levels 5 and 6 as low.

⁴ All personal information was collected by means of a survey questionnaire with informed consent.

TABLE 2
Personal Data of the Participants

Category	Subcategory	<i>n</i> (%)
Gender	Male	39 (50.0%)
	Female	39 (50.0%)
CSAT Level	Level 1	11 (14.1%)
	Level 2	10 (12.8%)
	Level 3	19 (24.4%)
	Level 4	18 (23.1%)
	Level 5	12 (15.4%)
	Level 6	3 (3.8%)
	Not Answered	5 (6.4%)
Self-rated Proficiency	Above Average	1 (1.3%)
	Average	39 (50.0%)
	Below Average	34 (43.6%)
	Not Answered	4 (5.1%)

The participants' CSAT levels (of English, hereinafter) show a relatively even distribution across Levels 1 through 6, out of the total 9 levels, and half of the participants considered themselves to be average in terms of English competence. Participants were also asked to provide another official test score, if any, for correlation examination with the CSAT. As only 11 participants had taken such a test, TOEIC,⁵ however, it was excluded from analysis.

2. Materials and Procedures

The reading text was an essay from Reading the World Now 3 (Jordens & Zeter, 2013), titled "Spreading the Seeds of Early Civilization," comprising 712 words in six paragraphs (see Appendix). The text readability level⁶ was assumed to be a little higher than the participants' current level. The topic of history was chosen as an area not familiar to the participants in order to minimize the background knowledge effect on reading

⁵ The scores ranged between 850 and 400.

⁶ It was the eleventh-grade level, according to The Flesch Grade Level Readability, available at <http://www.readabilityformulas.com/flesch-grade-level-readability-formula.php>.

comprehension and subsequent higher chances of picking up words due to content knowledge (Chen & Donin, 1997).

Glossing was selected as input modification to enhance the readability of the reading material and incidental vocabulary learning of participants. In fact, the participants were accustomed to such glossing as their current course book had right-column glosses for vocabulary. The total pool of target words was made up of those in the glossary of the essay and additional items of the researchers' judgement to make the reading less challenging, amounting to 30 words. The reading material was prepared with the thirty-word glossary in the right-hand column (see Appendix).

Following the first process of TMR, namely the CTI, participants completed a ten-minute PT with the thirty words. They were required to write the words' meanings in Korean before carrying out the reading activity.⁷ This previous exposure to the target words may have influenced the subsequent activities, but as all participants were in the same condition, it would not affect the bottom line of the research.

Following the PT, the participants completed a 30-minute reading activity, having been told that a multiple-choice test to assess their understanding of the text would follow; this was aimed at ensuring that they focused on reading comprehension. Instead, as planned, they completed a 10-minute multiple-choice vocabulary test for the thirty words after the reading activity. They were later told that this was done to ensure reliable results for the research. The DT was completed without warning, four weeks later in the same format as the IT but with a mixed order of multiple-choice items. Participants were informed that the activity and tests would count toward participation points for course grading in order to secure their full involvement. Every step of the procedure was closely supervised by the researchers.

3. Data Analysis

The three tests (PT, IT, and DT) containing thirty items each, were manually marked by the researchers. For the PT, which required the meanings of the words to be written in Korean, scoring standards were set up with a conservative approach. The following three cases were all counted correct and excluded from the target words: (a) for polysemes such as *domesticate*, *elaborate*, *pursuit*, *vessel*, and *cast*, answers of any one among their

⁷ The conventional SCS method, as described in III, typically involves a fixed number of target words for every participant. In this research, however, to make the results comparable, both SCS and TMR needed to be carried out in one experiment, in which CTI had already been applied. It should be noted that the SCS result in this part of the study is also based on the varying, customized number of target words, as in the case of the TMR, but the DT did not trace the IT.

multiple senses; (b) answers showing incorrect part-of-speech knowledge but correct meaning knowledge; and (c) answers that were not exactly correct but suggested understanding of the meaning.⁸ Following this, the results of the three tests were captured in an Excel worksheet with each item scored as 1 or 0, producing 30 cells for each test, making a perfect score 30 points. As the IT and DT were completed during different classes, the data of students who had missed either one were excluded from the analysis. As described above, the raw scores of the IT reflected only the cells of items that were answered incorrectly in the PT. The raw scores for the DT were calculated in two ways: for the SCS, by including incorrect items in the PT; and for the TMR, by including items answered correctly in the IT, using the IF AND formula, as illustrated in Section II.2.

IV. RESULTS

1. Effects of the Number of Target Words on SCS and TMR

Table 3 shows the descriptive statistics of the non-comparable⁹ raw scores of the three tests. The normalized and therefore comparable figures in Table 4 were calculated on the basis of the scores. This also includes two versions of the retention ratio: the SCS indicating the retention ratio based on the conventional measure, and the TMR indicating the ratio based on the new measure.

According to Table 4, applying conventional SCS, the mean scores for the IT and DT were 72.65 and 66.02, respectively, indicating a 6.63 percentage-point decline between the two tests; thus, the value of the variable SCS is 93.37. On the other hand, the DT in the TMR tracked the words answered correctly in the IT; thus, the mean score of the TMR indicates that participants remembered 77% of the words they initially acquired through the reading activity. A sizable difference exists between these two retention ratios. An intriguing result is that, in SCS, the maximum value is 115.00, suggesting that there were participants who performed better in the DT than in IT, which is unusual considering that humans tend to forget things over time.

⁸ Partial knowledge of word meaning can affect the results of the subsequent IT and DT as they are multiple-choice tests.

⁹ As the number of target words varies individually, raw scores of the IT that reflect the numbers of correct words out of the target words asked, for example, are not comparable each other.

TABLE 3
Descriptive Statistics: Raw Scores

	V1	V2	V3	V4
Mean	21.96	15.40	13.97	11.86
Median	23.00	15.50	13.50	11.00
Std. Deviation	6.01	4.24	3.72	3.87
Minimum	8.00	7.00	6.00	3.00
Maximum	29.00	25.00	23.00	20.00

V1: among 30 target words, # of words a participant answered incorrectly in pre-knowledge test

V2: among [V1], # of words a participant answered correctly in immediate test

V3: among [V1], # of words a participant answered correctly in delayed test in SCS

V4: among [V2], # of words a participant answered correctly in delayed test in TMR

TABLE 4
Descriptive Statistics: Normalized Scores (Percentages)

	Immediate Test ^a	Delayed Test for SCS ^b	SCS ^c	TMR ^d
Mean	72.65	66.02	93.37	77.00
Median	73.33	64.50	92.86	78.17
Std. Deviation	16.35	15.06	9.87	14.94
Minimum	34.48	24.14	70.37	33.00
Maximum	100.00	100.00	115.00	100.00

a: $V2/V1*100$

b: $V3/V1*100$

c: $100-(a-b)$: Conventional retention ratio

d: $V4/V2*100$: New retention ratio

Table 5 shows the results of a Pearson correlation analysis conducted to examine the relationship between the variables. They revealed the relative value of the SCS and TMR methods. Note that the cognitive load increases as the number of target words increases (see Sweller, 2005), as the work required for processing information becomes greater. Thus, heavy cognitive load leads to negative effects on task performance. However, Table 5 indicates that the SCS is not significantly correlated with the number of target words,¹⁰ whereas the TMR shows a negative correlation that is statistically significant at the .01 level, reflecting that a higher number of target words led to a lower TMR retention ratio.

¹⁰ The number of target words, hereinafter, is a variable indicating the number of unknown words, which a participant answered incorrectly, out of 30 words in PT.

TABLE 5
Pearson Correlations Between Variables ($n = 78$)

		Number of Target Words	SCS	TMR
Number of Target Words	Pearson	1.000	.056	-.440***
	Sig. (2-tailed)		.624	.000
SCS	Pearson	.056	1.000	.395***
	Sig. (2-tailed)	.624		.000
TMR	Pearson	-.440***	.395***	1.000
	Sig. (2-tailed)	.000	.000	

*** Correlation is significant at the .01 level (2-tailed).

As the result of correlation analysis supported the TMR over the SCS method, regression analyses were carried out to estimate how the SCS and TMR might change when the number of target words is varied. Panel A in Table 6 shows that the dependency of the SCS on the number of target words is not statistically significant ($F = .242, p = .624$). On the other hand, Panel B in Tables 6 and 7 indicate that TMR maintains a significant relationship with the number of target words ($F = 18.216, p = .000$) with the beta value of -.440.

TABLE 6
Regression: ANOVA

[Panel A] Dependent Variable: SCS; Predictors: (Constant), Number of Target Words					
Model	Sum of Squares	<i>df</i>	Mean Square	<i>F</i>	<i>Sig.</i>
Regression	23.862	1	23.862	.242	.624
Residual	7,478.907	76	98.407		
Total	7,502.769	77			
[Panel B] Dependent Variable: TMR; Predictors: (Constant), Number of Target Words					
Model	Sum of Squares	<i>df</i>	Mean Square	<i>F</i>	<i>Sig.</i>
Regression	.332	1	.332	18.216	.000
Residual	1.387	76	.018		
Total	1.719	77			

TABLE 7
Regression of TMR: Coefficients^a

	Unstandardized Coefficients		Standardized Coefficients	<i>t</i>	<i>Sig.</i>
	<i>B</i>	Std. Error	Beta		
(Constant)	1.011	.058		17.343	.000
# of Unknown Words	-.011	.003	-.440	-4.268	.000

a: Dependent Variable: TMR

2. Effects of CSAT Levels on SCS and TMR

A one-way ANOVA and post-hoc tests were conducted for the factors of gender, major, and self-rated proficiency, in addition to CSAT level, to see if those factors could provide any additional clues for the validation of TMR. Gender did not lead to any significant difference in the number of target words, i.e., unknown words in the PT, whereas all other factors did ($p = .000$). More precisely, participants' major, self-rated proficiency, and CSAT levels were all somehow related to their level of vocabulary. The latter two factors also showed significant group differences for TMR retention ratio, while the SCS did not show any meaningful differences in gender, major, self-rated proficiency, or CSAT levels ($p = .117, .839, .064, \text{ and } .112$, respectively).

It was also found that the group differences for major were caused by the Premed participants being mainly at Level 1 in the CSAT, and the ANOVA result for self-rated proficiency paralleled that of CSAT level. In short, all of those results signaled that the CSAT level was the most representative factor to vocabulary retention, and therefore can indicate the superiority of one model over the other.

Thus, the SCS and TMR versions of vocabulary retention ratio were examined as separate dependent variables along with the number of target words, with the focus on the CSAT level. The results of a one-way ANOVA, presented in Table 8, show that CSAT level led to significant differences in the number of target words ($F = 16.268, p = .000$) and TMR ($F = 6.976, p = .000$), but not in the case of SCS ($F = .337, p = .888$). It seems reasonable that higher performing learners have smaller numbers of unknown words (target words) than those of lower performance. It is also reasonable that the vocabulary retention ratio is higher among those of higher performance, as they have fewer target words, generating a lower cognitive load. Thus, the fact that CSAT level does not have a significant impact on SCS can serve as a very practical support for TMR.

TABLE 8
Effect of CSAT on Dependent Variables: One-way ANOVA

Dependent Variables		Sum of Squares	<i>df</i>	Mean Square	<i>F</i>	<i>Sig.</i>
Number of Target Words	Between Groups	1,478.179	5	295.636	16.258	.000
	Within Groups	1,218.342	67	18.184		
	Total	2,696.521	72			
SCS	Between Groups	162.886	5	32.577	.337	.888
	Within Groups	6,468.385	67	96.543		
	Total	6,631.270	72			
TMR	Between Groups	.577	5	.115	6.976	.000
	Within Groups	1.108	67	.017		
	Total	1.685	72			

Table 9 shows the results of post-hoc analyses making multiple comparisons across CSAT levels. The focus was on Levels 3 and 4, as the majority of participants fell into these categories ($n = 19$ and 18 , respectively, and 47.44% in total). Two noteworthy findings are clear from Table 9. First, there were no significant differences revealed between Levels 3 or 4 and any of the others in terms of the retention ratio under SCS, which may be taken as evidence that SCS is less reliable. Second, whereas Levels 3 and 4 are jointly considered as the middle proficiency group at the university, there seems to be a fine dividing line between the two. For example, Level 4 differed significantly from Levels 1 and 2 in terms of TMR ($p = .005$ and $.003$, respectively), but Level 3 did not; this may suggest potential homogeneity across the first three levels, which is a finding that may potentially inform class placement at the university where the research was carried out.

TABLE 9
Multiple Comparisons for CSAT Levels: Post Hoc Tests (Bonferroni)

Dependent Variables	(I)SAT Level	(J)SAT Level	Mean Diff. (I-J)	Std. Error	Sig.	95% Confidence Interval		
						Lower Bound	Upper Bound	
Number of Target Words	3	1	9.019*	1.616	.000	4.10	13.94	
		2	4.774	1.666	.083	-.30	9.85	
		4	-.971	1.403	1.000	-5.24	3.30	
		5	-4.693	1.572	.059	-9.48	.09	
		6	-6.193	2.649	.336	-14.26	1.87	
		4	1	9.990*	1.632	.000	5.02	14.96
	2	5.744*	1.682	.016	.62	10.86		
	3	.971	1.403	1.000	-3.30	5.24		
	5	-3.722	1.589	.332	-8.56	1.12		
	6	-5.222	2.659	.806	-13.32	2.87		
	SCS	3	1	1.219	3.723	1.000	-10.113	12.551
			2	-1.834	3.839	1.000	-13.519	9.852
4			2.562	3.232	1.000	-7.276	12.400	
5			-.526	3.623	1.000	-11.555	10.503	
6			2.249	6.104	1.000	-16.333	20.831	
4			1	-1.343	3.760	1.000	-12.790	10.104
2		-4.396	3.875	1.000	-16.193	7.401		
3		-2.562	3.232	1.000	-12.400	7.276		
5		-3.088	3.662	1.000	-14.235	8.059		
6		-.313	6.127	1.000	-18.966	18.339		
TMR		3	1	-.059	.049	1.000	-.207	.090
			2	-.072	.050	1.000	-.225	.081
	4		.128	.042	.054	-.001	.256	
	5		.135	.047	.086	-.009	.280	
	6		.190	.080	.306	-.054	.433	
	4		1	-.187*	.049	.005	-.336	-.037
	2	-.199*	.051	.003	-.354	-.045		
	3	-.128	.042	.054	-.256	.001		
	5	.008	.048	1.000	-.138	.154		
	6	.062	.080	1.000	-.182	.306		

* The mean difference is significant at the 0.05 level.

V. DISCUSSION

The quantitative research of this paper was designed to demonstrate the adequacy of the newly proposed TMR model by means of an empirical examination of vocabulary development. As most studies in the similar vein have applied the conventional SCS as previously reviewed, we employed both models of measurement for an experiment to compare and evaluate their plausibility. The results were in support of the TMR, which can be recapitulated by answering the second and third research questions: *Do the SCS and TMR models of measurement yield different results for vocabulary retention ratio, and, if so, which of the models is more accurate?*

The two models did indeed produce different results, highlighting three points worthy of discussion. First, the maximum of the normalized SCS, i.e., the retention ratio calculated in the conventional SCS, was 115; the literal interpretation would be that some students recalled words that they had not learned previously. The maximum TMR, in contrast, was 100, meaning that some students recalled every word they had learned previously. Second, in the correlation analysis between variables, the number of target words correlated significantly with TMR but not with SCS. The target words are the unknown words identified through the PT, the number of which varies for each participant. It is highly likely that the retention ratio decreases when more words are unknown, due to increased cognitive load. Third, the post-hoc between-group comparisons of CSAT levels versus TMR generated more likely results. For example, Levels 1 and 2 participants maintained a significantly higher retention ratio than did Level 4 participants. On the other hand, SCS did not show any significant differences between any pairs of CSAT levels. All these results turned out to flow in the same direction, the better plausibility of TMR.

Consider again the first research question: *Does CSAT level significantly influence learners' vocabulary development (retention) following an in-class reading activity?* The present findings suggest a positive answer to the question by means of the above-mentioned findings that corroborated TMR as a more accurate and precise measure of vocabulary retention. Thus, in addition to the fulfillment of the original objective of this study, to verify the TMR model, the findings of this empirical study may be considered as further confirmation of those of previous studies that have reported a positive impact of proficiency level on vocabulary learning.

VI. SUMMARY AND CONCLUSIONS

This paper has proposed a new model to measure L2 vocabulary retention for pedagogical research that may be a better index of learners' development in two ways.

First, it showed the conventional method, SCS, has the potential to incorrectly include items not actually retained, namely W10, as visualized in Figure 2, and suggested an alternative, the TMR. Second, to demonstrate the greater accuracy of the new TMR model over the SCS, the latter quantitative analysis of this paper reported on a study that applied both methods to an examination of vocabulary development following an in-class reading activity. The proficiency level has been shown by previous studies to have an impact on learners' incidental learning of vocabulary, and was therefore used here as a verification tool. The two types of retention ratios were compared against participants' proficiency level, and the results supported TMR over SCS.

The limitations of TMR may be self-evident. Its binary system optimizes multiple-choice tests, so additional processes may be necessary to accommodate open-ended questions allowing partial credit. In addition, it may not be for paper-based testing (PBT), and researchers may prefer to administer computer-based testing (CBT); because TMR requires data entry for each item, the amount of keyboarding may be excessive in case of PBT as the number of participants and target words increase.

The greatest advantage of TMR that can outweigh the inconvenience is that it successfully eradicates the necessity of assuming that all participants do not know a certain group of words, i.e., the assumption that an experimental group is homogeneous. This often reduces the impact of findings in this type of pedagogical research on L2 vocabulary learning. The present data actually showed that participants of the same CSAT level vary widely in terms of the number of unknown words out of the total 30 words. However precise a target word screening process is, it cannot reflect participants' individual variation in pre-knowledge enough as far as it relies on participants' declaration or prior investigation on a quasi-group. Researchers would benefit from customizing target words per participant in such studies.

A second great advantage of the TMR model is that it provides researchers with a useful and practical tool to measure the retention of vocabulary. TMR, which is a sufficiently more accurate measurement than simply measuring changes in scores, requires only Microsoft Excel, to capture the primary data and to use as a result compiling platform throughout the entire process of tracing PT to IT, and IT to DT.

In a nutshell, TMR deals with measuring the retention of previously learned knowledge as a true reflection of development. Its contribution may be small but fundamental in terms of evaluating L2 vocabulary development, and possibly in further pedagogical research areas as well. It is hoped that this study has provided future researchers with a means to measure learning precisely, which may usher them to insightful educational findings.

REFERENCES

- Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language, 20*(2), 136-163.
- Chen, Q., & Donin, J. (1997). Discourse processing of first and second language biology texts: Effects of language proficiency and domain-specific knowledge. *Modern Language Journal, 81*(2), 209-227.
- Coady, J. (1997). L2 vocabulary acquisition through extensive reading. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 225-237). New York: Cambridge University Press.
- Cunningham, A. C., & Stanovich, K. E. (2001). What reading does for the mind. *Journal of Direct Instruction, 1*(2), 137-149.
- Frishkoff, G. A., Perfetti, C. A., & Collins-Thompson, K. (2011). Predicting robust vocabulary growth from measures of incremental learning. *Scientific Studies of Reading, 15*(1), 71-91.
- Hayes, D. P. (1988). Speaking and writing: Distinct patterns of word choice. *Journal of Memory and Language, 27*(5), 572-585.
- Hemmati, T., & Asmawi, A. B. (2015). Incidental vocabulary learning and retention through reading a graded reader among Iranian EFL learners. *The Online Journal of New Horizons in Education, 5*(1), 72-86.
- Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language. *Studies in Second Language Acquisition, 21*(2), 181-193.
- Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *Modern Language Journal, 80*(3), 327-339.
- Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning, 51*(3), 539-558.
- Hunt, A., & Beglar, D. (2002). Current research and practice in teaching vocabulary. In J. C. Richards & W. A. Renandya (Eds.), *Methodology in language teaching: An anthology of current practice* (pp. 258-266). Cambridge: Cambridge University Press.
- Jenkins, J. R., Stein, M. L., & Wysocki, K. (1984). Learning vocabulary through reading. *American Educational Research Journal, 21*(4), 767-787.
- Jeong, Hogeong, Biswas, G., Johnson, J., & Howard, L. (2010). Analysis of productive learning behaviors in a structured inquiry cycle using hidden Markov models. *Proceedings of the Educational Data Mining 2010, Pittsburgh, USA, 81-90*.

- Jordens, R., & Zeter, K. (2013). *Reading the world now 3*. Seoul: Compass Publishing.
- Kim, Hyung-Sun, & Na, Yoon-Hee. (2010). Vocabulary learning and task-induced involvement. *Korean Journal of Applied Linguistics*, 26(4), 183-211.
- Kim, YouJin. (2008). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning*, 58(2), 285-325.
- Ko, Myong Hee. (2012). Glossing and second language vocabulary learning. *TESOL Quarterly*, 46(1), 56-79.
- Larsen-Freeman, D., & Long, M. H. (1991). *An introduction to second language acquisition research*. New York: Longman.
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22(1), 1-26.
- Lee, Shinwoong. (2013). Corpus-based vocabulary analysis of *Friends*: Implication for incidental vocabulary learning. *Modern English Education*, 14(3), 1-21.
- Nemati, A. (2013). Vocabulary learning strategies: A short way to long term retention. *Linguistic and Literature Studies*, 1(1), 8-14.
- Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 19-30). Cambridge: Cambridge University Press.
- Tekmen, E. A. F., & Daloglu, A. (2006). An investigation of incidental vocabulary acquisition in relation to learner proficiency level and word frequency. *Foreign Language Annals*, 39(2), 220-243.
- Yoshii, M., & Flaitz, J. (2002). Second language incidental vocabulary retention: The effect of text and picture annotation types. *CALICO Journal*, 20(1), 33-58.
- Yuksel, H. G., & Kavanoz, S. (2010). Does vocabulary knowledge distinguish among proficiency levels of Turkish university students? *The International Journal of Learning*, 17(5), 513-521.

APPENDIX

Reading Activity: Spreading the Seeds of Early Civilizations

<p>The first people to inhabit the earth lived more than two million years ago. They survived by hunting, fishing, and gathering wild plants. They were forced to move from place to place in the pursuit of food. The development of agriculture around 10,000 BC ended this nomadic way of life and gave rise to the first civilizations. When people learned how to grow crops and domesticate animals, they were able to settle in one place. Settlements grew into villages and eventually into cities. People began to work specialized jobs, develop</p>	<p>inhabit: 서식하다 pursuit: 추구 nomadic: 유목의 crop: 농작물 domesticate: 사육하다 emerge: 모습을</p>
--	---

<p>arts and crafts, and participate in other activities of civilized life. They formed governments and invented systems of writing. By 3500 BC, the world's first civilizations emerged in the fertile valleys of the Tigris-Euphrates, Nile, Indus, and Huang He rivers.</p>	<p>드러내다 fertile: 비옥한</p>
<p>The people who lived in the southern Mesopotamian Valley, between the Tigris and Euphrates river, were known as Sumerians. They first settled on small farms and built dikes and irrigation canals to control flooding and water their fields. The Sumerian villages soon grew into small cities. Many of these became powerful city-states by about 3200 BC. One of the greatest achievements of the Sumerian civilization was the development of a writing form known as cuneiform, which used wedge-shaped pictures for alphabetical characters. By carving these symbols into clay tablets, the Sumerians wrote letters as well as literary and religious texts, recorded legal and economic transactions, and studied mathematics, astronomy, and medicine. The Sumerians were fine artisans who produced board games, jewelry, metalware, glassware, musical instruments, and decorative pottery. The Sumerian city-states, however, lacked a centralized government and unified army to defend their civilization against invaders, and thus by about 2000 BC their civilization had fallen under Semitic rule.</p>	<p>dike: 제방 irrigation: 관개 canal: 수로 cuneiform: 설형문자 wedge-shaped: 췌기(ㄴ자) 모양의 artisan: 장인</p>
<p>The ancient peoples of the Nile River Valley in Egypt began developing their civilization around 3100 BC. The ancient Egyptians invented their own system of writing known as hieroglyphics. Many of the material accomplishments of ancient Egyptian society centered around religion. They performed elaborate burial procedures to assure that the dead were comfortable in the afterlife. They embalmed and dried the corpses in a process called mummification in order to preserve the bodies. They also built gigantic burial tombs that they filled with clothing, food, furnishings, and jewelry. For deceased kings, they built elaborate pyramids, which required advanced engineering knowledge and thousands of workers. Eventually, the Egyptians developed an army and had a fleet of trade ships at their disposal, expanding their kingdom and acquiring precious materials from all over the ancient world.</p>	<p>hieroglyphic: 상형문자 elaborate: 정교한 embalm: 방부처리를 하다 corpse: 시체 gigantic: 거대한 deceased: 죽은</p>
<p>In the Indus River Valley, two cities developed, which the Indus people called Mohenjodaro and Harappa. By 2500 BC, these cities together supported nearly 70,000 inhabitants, whose advanced system of agriculture was able to feed the large population. Indus cities were constructed primarily using bricks, and city streets were well planned in a rectangular grid pattern. The only traces of Indus writing were discovered on stone seals, which were engraved with animal and human forms as well as inscriptions that remain to be fully translated. The Indus people produced other crafts, including furniture, decorative</p>	<p>rectangular: 직사각형의 grid: 격자무늬 inscription: 새겨진 글 utensil: (가정용) 도구</p>

jewelry, metal utensils , and toys, which they traded with other civilizations as distant as Mesopotamia. Around 1700 BC, however, the Indus Valley civilization mysteriously disappeared.	
The decline of civilization in the Indus Valley marked the beginning of the Shang Dynasty in the valley of the Huang He. The Shang people developed a writing system using more than 3,000 characters which they scratched on animal bones and turtle shells. These bones and shells are known as oracle bones. Although they were primarily used in religious ceremonies to predict the future and solve worldly concerns, the oracle bones also provide a record of the people of the Shang Dynasty and events occurring during the Shang period. Shang cities were large and surrounded by high walls. The people cast bronze vessels , carved marble and jade, and wove silk. Shang kings organized large armies and ruled the Huang He Valley for over 600 years.	oracle: 신탁, 제사장 <i>oracle bone script</i> 갑골문 worldly: 세속적인 cast: 주조하다 vessel: 그릇 weave: 짜다
As the world's first civilizations advanced and spread, people began to exchange tools and ideas. Some customs and languages merged , while many others diverged and formed new cultures. The societies that would one day populate the world had begun to emerge.	merge: 융합되다 diverge: 나뉘다

Adapted from Jordens & Zeter (2013, pp. 54-55)

예시언어(Examples in): English

적용가능 언어(Applicable Languages): English

적용가능 수준(Applicable Levels): Tertiary

Hyung-Sun Kim
Division of Undeclared Majors
Chosun University
309 Pilmun-daero, Dong-gu, Gwangju 61452
Tel: (062) 230-6948
Email: hskim4@chosun.ac.kr

Ung-Yong Choi
College of Business Administration
Chonnam National University
77 Yongbong-ro, Buk-gu, Gwangju 61186
Tel: (062) 530-1476
Email: uychoi@jnu.ac.kr

Received 15 December 2016

Reviewed 20 January 2017

Accepted 10 February 2017