영어어문교육 14권 1호 2008년 봄

Validation of the Common Metric Scale (CMS)

Yong Myeong Kim
(Hankuk University of Foreign Studies)

Kim, Yong Myeong. (2008). Validation of the common metric scale (CMS). *English Language & Literature Teaching*, *14*(1), 21-44.

On the basis of the Parallel Developmental Sequence (PDS) Model (Kim & Kwon, 2005, 2007), Kim (2006) proposed the *Common Metric Scale (CMS) Model* for measuring language ability. The CMS was developed in terms of the Interlanguage (IL) approach, the Testability, and the Bounded-domain hypothesis. The IL approach enables us to map the PDS onto the CMS. Thus, the absolute scales, the scale calibration, and the scale descriptions of the CMS can be automatically defined. In addition, the Testability hypothesis enables us to gear test tasks to the Parallel Stages of the PDS, and the Bounded-domain hypothesis enables us to synchronize test domains with IL discourse domains. Thus, the resulting CMS satisfies what we call the *triple implicational scales* in which the scales, the scale descriptors, and the task hierarchies are all in an implicational relationship. To validate the theoretical CMS, the Spearman Correlation and the Kruskal-Wallis Test were conducted. The results demonstrated that the CMS has its own validity. In sum, the CMS based on the PDS can function as an *IL-sensitive scale* for identifying IL learners' current developmental stages, diagnosing the deficiencies in their IL grammar, and predicting what rules or structures will be learnable or teachable next, thereby giving them a "timely remedy and tailored teaching".

[SLA/parallel developmental sequence/common metric scale/IL-sensitive scale]

I. INTRODUCTION

As a future direction toward language testing, Bachman (1990) proposed the development of a *common metric scale*[1] for measuring language abilities in a wide variety of contexts, at all levels, and in many different languages in order to meet various

[1] Clark (1980, p. 15) defined a common metric scale as "a uniform testing procedure that could be used with diverse groups of examinees in a variety of language learning and language-use situations, with testing results reported on a uniform scale" (recited from Bachman, 1990, pp. 339-340).

needs for language teaching, language testing, and second language acquisition (SLA) research. In addition, since the late 1980's, several researchers have increasingly recognized that there should be an *interface* between SLA research and language testing (Bachman, 1990; Bachman & Cohen, 1998; Brindley, 1998; Douglas, 1998; Long & Richards, 1998; Shohamy, 1998). Reviewing this research has come to a convergence on *reciprocal developments* between these two disciplines in that language testing both serves and is served by SLA research (Bachman, 1990). More specifically, as Shohamy (1998) mentioned, developments in language testing can provide SLA research with psychometric criteria for evaluating tasks and instruments to ensure that the information obtained from these tasks is reliable and valid. Conversely, as Long and Richards (1998) suggested, advances in SLA can throw new light on such crucial notions for test developers as acquisition and developmental sequence, and on the relevant criteria for student placement for the purpose of instruction, leading in some quarters to requests for tests that are sensitive to intelanguage (IL) development, and that take the individual learners as the unit of analysis. Then, where is a *parallel interface* to enhance reciprocal developments between SLA research and language testing?

In recent years, from the reviews of such SLA theories as the Multidimensional Model (Meisel, Clahsen, & Pienemann, 1981; Pienemann & Johnston, 1987), the Minimal Tree Approach (Kim Boram, 2006; Park, 2006; Vainikka & Young-Scholten, 1998), and the Processability Theory (Pienemann, 1998), Kim and Kwon (2005, 2007) proposed a modular model called *the Parallel Developmental Sequence (PDS) Model.* The PDS comprises the Parallel Developmental Stages implicationally called $1X,\cdots, 1X^2,\cdots, 1X^3,\cdots, kX^n$ (here, n is the number of embedded clauses). Each parallel stage of the PDS incorporates three different dimensions of sequential developments: the Procedural, the Syntactic, and the Morphological Sequence. Thus, the PDS provides an interface between the three sub-developmental sequences (for full discussion of the PDS Model, see Kim & Kwon, 2007). In sum, the PDS Model is a kind of acquisitional map along which L2 learners assume to proceed from the initial stage to the next higher stage, in a stepwise fashion, until up to the final stage on the PDS. Just as when we are on our trip to a strange place, by relying on the map of that area, we can find out our journey route from where we are now to where we are going to go next, so, by analyzing L2 learners' IL data, we can find out not only where they are now on the acquisitional map, that is, what structures or rules they are able to process, parse, and unify in terms of the PDS, but also we can predict where they are going to go next on the map, that is, what structures or rules they are able to acquire next in terms of the PDS. Therefore, the PDS Model can provide a theoretical basis for developing an *acquisition-oriented* or an *IL-sensitive scale* that are sensitive to, or tailored for language developmental sequence.

On the basis of the PDS Model, Kim (2006) proposed a *Common Metric Scale (CMS) Model*

for measuring language ability in order to constitute a parallel interface between SLA research and language testing. The CMS was developed in terms of the IL approach which converts the PDS into the CMS, and hence the absolute scales, the scale calibration, and the scale descriptions of the CMS can be automatically defined. The resulting CMS comprises three componential or analytic scales, that is, the Syntactic Scale, the Procedural Scale, and the Morphological Scale, with each scale corresponding to each developmental sequence of the PDS, that is, the Syntactic, the Procedural, and the Morphological Sequence. Also, the CMS consists of a total of 10 scales from the absolute "zero" proficiency (Level 0) to the absolute "perfect" proficiency (Level 9) (Full discussion will be elaborated on in the next section). In sum, not only can the CMS based on the PDS constitute a parallel interfaces between SLA research and language testing, but it also can function as an *IL-sensitive* or an *acquisition-oriented scale* for implementing a "timely remedy and tailored teaching".

The question then is whether or not the theoretically built CMS does its validity. Thus, to see whether there is a relationship between the ratings on the CMS and the scores on a standardized test (i.e., the Grammar test and the Listening Comprehension test of the TEPS), the Spearman Correlation and the Kruskal-Wallis test will be conducted. Finally, the results and implications will be discussed, relating to those of Kim's research (2006) into the relationship of the CMS and the proficiency of Grammar and Reading Comprehension..

II. THEORETICAL BACKGROUND

1. The Real-Life (RL) and the Interactional/Ability (IA) Approach

Bachman (1990) drew a distinction between the real-life (RL) and the interactional/ ability (IA) approach to defining language ability and test authenticity. In the RL approach, since language proficiency is viewed as unitary ability, and defined in terms of language performance, there is no basis for distinguishing separate abilities such as grammatical, pragmatic, and sociolinguistic competence. Thus, as is the case with the ACTFL guidelines, separate abilities cannot be measured as a part of language proficiency, and hence scores are expressed in a single or global rating of general language ability. Also, since test authenticity is defined in terms of "real-life" language performance, there is no distinction between the ability to be measured and the features of the context in which the performance takes place. Thus, scale definitions include facets of the test elicitation procedures. Although the behavioral scales based on the RL approach such as the ACTFL guidelines and ILR scales have had strong influences on testing practice until now, the validity of these scales as indicators of language ability has been challenged because the scales fail to distinguish between the ability being assessed and the elicitation procedures

used, thus limiting the generalization of scores to non-test situations (Bachman, 1990). The usefulness of the scales as diagnostic and achievement measures has also been questioned because they fail to provide a "profile" of each component of language abilities (Bachman & Palmer, 1996). On the other hand, in the IA approach, as is the case with the Bachman and Palmer Scales (Bachman & Palmer, 1983), since we define language proficiency in terms of components of abilities, we should attempt to measure these component abilities and report scores on an analytic scale. Similarly, since we define test authenticity in terms of the interaction between the abilities of the language user and the features of the language use context, scale definitions of abilities necessarily exclude facets of the test methods.

Based on the IA approach, Bachman (1990) and Bachman and Palmer (1996) presented an approach to the development of a *common metric scale* (CMS). According to Bachman and Palmer (1996), the approach includes two principles. One principle is that the operational definitions of the scales are based on the theoretical definitions of the constructs, that is, the componential definitions of language abilities such as communicative language ability (Bachman, 1990). Following this principle, they developed analytic or componential scales, which require the rater to provide separate ratings for the different components of language abilities in the construct definitions. The other principle is that the scales are defined operationally in terms of criterion levels of ability, rather than as a performance relative to other test takers or to native speakers of the language. Bachman (1990) suggested criterion-referenced scales, with criteria consisting of absolute levels, which may or may not be found in actual language users, defined in terms of each component of language proficiency.

However, from an SLA perspective, as Kim (2006) pointed out, the IA approach can be challenged in that it fails to relate developmental information and insights from sequence-based SLA research such as the Multidimensional Model (Pienemann & Johnston, 1987), and the Processability Theory (Pienemann, 1998) to the developments and constructions of a CMS, thus having a limitation in helping to enhance test authenticity. Scales based on the IA approach can also be questioned in that they fail to reflect the IL developmental sequence from sequence-based SLA research, thus having a problem in functioning as an IL-sensitive measure for assessing and diagnosing IL learners' language proficiency.

2. The Interlanguage (IL) Approach to a Common Metric Scale

As a possible complement to the IA approach, Kim (2006) proposed the interlanguage (IL) approach, the Testability and the Bounded-domain hypothesis, which are a kind of "converters" that convert the PDS into an *IL-sensitive* or *acquisition-oriented scale*.

The IL approach was derived from the developmental approaches (Ingram, 1985;

Pienemann & Mackey, 1993). It relates the developmental stages and the governing mechanisms of the PDS to the overall plans for developing and implementing a CMS. The IL approach thus enables us to map the PDS onto a CMS. The Testability hypothesis was extrapolated from what Pienemann calls the Teachability principle (1985, 1998). Given that any instruction will be beneficial when it focuses on structures or rules of the next or subsequent stage of the IL learners' current stage on the developmental sequence, then, any test task will be "authentic" to the test taker if it relates to structures or rules of the next or subsequent stage of the IL learners' current stage. Thus, the Testability hypothesis enables us to gear test tasks or items to IL learners' language abilities. The Bounded-domain hypothesis was derived from intertwining what Douglas and Selinker (1985) call the *discourse domains* and what Pienemann (1998) calls the *Steady hypothesis*. Douglas and Selinker (1985) presented data from both IL research data and speaking tests to support one of their testing principles that "test takers create individual discourse domains, which are a function of the discourse domains they have already created in their IL use" (recited from Bachman, 1990, p. 324). They also presented another principle that "the closer the context supplied by the test writer to prototypical internal IL contexts created by the test taker, the more likely it is that the test will engage the test taker's ability to perform the task at hand in the L2" (p. 220). Pienemann (1998) undertook task variation research in which ESL learners were required to carry out different communicative tasks. On the analysis of the IL profiles of the learners, he proposed the Steady hypothesis, according to which the basic nature of the grammatical system of an IL does not change in the different communicative tasks. What both studies imply is that content or task variation resulting from test performances is not a "free" but a "bound" variation. Thus, if we combine Douglas and Selinker's discourse domains with Pienemann's Steady hypothesis, then, an interesting hypothesis will emerge, namely, that discourse domains will extend not so much "unboundedly", unshackled as "boundedly", interlocked with syntactic developments. The hypothesis hence postulates that as IL learners proceed from one stage to the next higher stage on the PDS, so will they create and extend their own IL discourse domains along a continuum ranging from more concreteness to more abstractness. Both IL discourse domains and test domains are, in general, assumed to constitute a continuum of the IL discourse domains ranging from everyday situations and contexts (e.g., greeting) to more specific ones (e.g., court), and from more general or concrete topics and contents (e.g., food) to more abstract or academic ones (e.g., philosophy). Thus, the Bounded-domain hypothesis enables us to synchronize test domains with the IL discourse domains created and expanded by IL learners. In conclusion, not only can the IL approach offer a possible solution to the limitations inherent in the RL or the IA approach, but it also can map the PDS onto a CMS, and thus the resulting CMS can be an *IL-sensitive* or an *acquisition-oriented scale.*

3. The Theoretical Common Metric Scale (CMS) on the IL Approach

From the reviews of language testing studies such as Bachman (1990), Bachman and Palmer (1996), and Brindley (1998), Kim (2006) identified four problems to be solved in developing a CMS as follows; (1) how to define absolute levels on what basis, (2) how many scales to have between the absolute levels, (3) how to specify the scale definitions, and (4) how to establish an implicational relationship between the scale descriptions and the hierarchies of task difficulty. On the discussions of these four problems in terms of the IL approach, Kim (2006) proposed the *Common Metric Scale (CMS)* based on the PDS as illustrated in Table 1. Recapitulate on how the IL approach deals with these problems, and thus how it maps the PDS onto the CMS.

TABLE 1

The Common Metric Scale (CMS) for Language Proficiency on the PDS (Kim, 2006, p. 91)

Parallel Stages	Parallel Developmental Sequence (PDS)			Critical Structures	Error Patterns	CMS Levels
	Syntactic Scale	Procedural Scale	Morphological Scale			
	Parallel Mechanisms					
	Syntactic M	Procedural M.	Unificational M.			
1X	Lexical-inserting	[· IFM][+COM]	No unification	Canonical SVO	…	**0**
2X	Theta-marking	[+IFM] [+COM]	Phrasal [$_{VP}$…]	*Wh/Do*-frontings	…	**1**
3X	Case-marking	[+IFM][-COM]	Clausal [$_{IP}$…]	*Aux-en/-ing*, Y/N-questions	…	**2**
4X	Force-realizing	[-IFM][-COM]	Sentential [$_{CP}$…]	*Wh*-questions, *Neg*-Inv.	…	**3**
1X²		[· SCM([· IFM][+COM])]		2-sub comp, *that/for*-clauses	…	**4**
2X²		[+SCM([+IFM][+COM])]		Indirect-questions Relatives (subject)	…	**5**
3X²	Force-matching, Markedness on Operation	[+SCM([+IFM][-COM])]	Intersentential [$_{CP}$ [$_{CP}$…]]	Relatives (object, oblique)	…	**6**
4X²		[+SCM([-IFM][-COM])]		Indirect-questions (object)	…	**7**
5X²		[-SCM([-IFM][-COM])]		Long distance-questions Cleft sentences	…	**8**
***k*X³**	…	[±SCM([±SCM])]	…	Double complex sentences	…	**9**

1) On What Basis Can We Define Absolute Levels?

The CMS on the IL approach is based on a theory of language acquisition, more precisely, the PDS Model, thus constituting an interface between SLA research and language testing. As seen in Table 1, the CMS on the IL approach consists of three

componential or analytic scales, the Syntactic Scale, the Procedural Scale, and the Morphological Scale, with each scale corresponding to each sequence of the PDS, that is, the Syntactic, the Procedural, and the Morphological Sequence. As Bachman and Palmer (1996) mentioned, the use of analytic scales can provide a "profile" of the components of language ability by differentiating among areas of relative strength and weakness.

Related to this problem is the question of how to define the absolute levels, "zero" at the bottom and "perfect" at the top. Bachman (1990) argued against defining *an educated native speaker* as the criterion for perfect proficiency (based on the RL approach) since any scale based on actual native speakers' performances becomes, by definition, a norm-referenced scale. He, instead, defined absolute levels, which are essentially *hypothetical levels* that function as a basis for scale definitions and calibration, but which may or may not be found in actual language users (based on the IA approach). However, even if absolute levels are defined as hypothetical levels, there still remains a problem with the definition in that they can not be defined concretely. In a sense, "hypothetical levels" might be merely an embellishing of "an educated native speaker" with "hypothetical".

On the other hand, following the IL approach, the absolute scales of the CMS can be automatically defined in terms of the lowest (initial) and the highest (final) stage of the PDS. Thus, Stage 1X of the PDS is equal to the absolute "zero" proficiency (Level 0), and Stage kX^3 to the absolute "perfect" proficiency (Level 9) of the CMS as seen in Table 1.

2) How Many Scales Should There Be Between the Absolute Levels?

Bachman (1990) regarded this problem as an empirical one, and proposed that the number of intermediate levels is determined through a rather complicated program of research and development. In such a program, initially, the absolute levels (zero and perfect proficiency) are defined, and then several intermediate levels are provisionally posited on the ability continuum between two end points. Next, a series of elicitation procedures is designed to be applicable to several different learner groups ranging from "novices" with little or no ability to "experts" with a high level of ability. Then, the authentic samples elicited by the procedures are rated on the ability scales. Finally, the calibration of both the abilities and the scale points is analyzed and revised.

Yet, to follow the IL approach would result in the same results as Bachman's approach. That is, the number of intermediate levels is determined automatically but empirically in terms of the number of Parallel Stages in the PDS, thereby the resulting CMS consisting of a total of 10 scales (scale 0-9) as seen in Table 1[2].

[2] In fact, the CMS has a degree of flexibility in that the number of scales can be tailored depending on the levels of test takers' proficiencies (e.g., beginner or advanced learners), the types of test tasks

3) How Can We Specify the Scale Descriptions?

With regard to the scale descriptions, Brindley (1986, 1998) criticized behavioral rating scales such as the ACTFL guidelines because their scale descriptors are described by relative "impressionistic terms" such as "fluent", "intelligible", "most", "some", and thus subjective judgment would have to play a role, thereby clouding test reliability. Bachman (1990) argued that since scale descriptions based on the RL approach are norm-referenced, and do not distinguish the components of proficiency from test method facets, they may not be adequate for making inferences about how much language ability a test taker has, and about how score interpretations generalize beyond performance on the test to other non-test situations. He advocated criterion-referenced rating scales based on the IA approach such as the Bachman-Palmer Scales (Bachman & Palmer 1983) as seen in Table 2.

TABLE 2

Bachman-Palmer Scales: Grammatical Competence (Bachman, 1990, p. 327)

Rating	Range	Accuracy
0	*No systematic* evidence of morphologic and syntactic structures	Control of *few or no* structures; errors of *all or most* possible types
1	*Limited* range of both morphologic and syntactic structures, but with some systematic evidence	Control of *few* or no structures; errors of *all or most* possible types
2	*Limited* range of both morphologic and syntactic structures, but with some systematic evidence	Control of *some* structures used, but with *many* error types
3	*Large, but not complete*, range of both morphologic and syntactic structures	Control of *some* structures used, but with *many* error types
4	*Large, but not complete*, range of both morphologic and syntactic structures	Control of *most* structures used, with *few* error types
5	*Complete* range of morphologic and syntactic structures	Control of *most* structures used, with *few* error types
6	*Complete* range of morphologic and syntactic structures	No systematic errors

The scale descriptors on the IA approach are, in fact, more advanced than ones on the RL approach in that they have been developed through such a complex program as mentioned above, but they seem to run up against the same problem as the behavioral descriptors. As seen in Table 2, for example, "relative terms" such as "*complete*", "*some*"

(e.g., oral interviews or multiple choice tests), and the purposes of testing (e.g., diagnostic tests or placement tests). For example, in case test takers are novice, say, elementary students, we can incorporate all the levels of Stage X^2 into a single highest level, that is, level 4, leaving the levels of Stage X intact. The resulting CMS consists of a total of 5 levels from level 0 to 4. On the other hand, in case they are advanced learners, say, college students, we can incorporate all the levels of Stage X into a single lowest level, that is, level 0, keeping the levels of Stage X^2 intact. The resulting CMS comprises a total of 6 levels from level 0 to 5. In order to present a "prototypical" scale, this study defined the scale of the CMS as 10 levels.

and "*many*" are still used in the Bachman-Palmer Scales, and indeed, "how *many* is *many*?" In addition, as Brindley (1998) pointed out, how could negatively defined descriptions, for example, "*Large, but not complete*" as in Rating 3 of the Table 2, be translated directly into test specifications or test items?

According to the IL approach, the scale descriptions of each level can be described both consistently and sequentially by the implicational scale inherent in the PDS. That is, as seen in Table 1, the descriptor for Level 0 corresponds to the parallel mechanism, critical structures, and error patterns of Parallel Stage 1X, and the descriptor for Level 1 to those of Parallel Stage 2X on the PDS, and so on. According to Kim and Kwon (2007), the PDS constitutes an implicational scale in which the higher stage can subsume the parallel mechanisms, the critical structures and error patterns of the lower stages, but not *vice versa*. Thus, the descriptors of each level of the CMS are also in an implicational relationship. In sum, not only can the IL approach offer a possible solution to the problems of the scale descriptors on the RL or the IA approach (i.e., imprecision or inconsistency resulting from using impressionistic, subjective, or relative terms), but it also can constitute an implicational relationship between scales and scale descriptors. To compare Bachman-Palmer's Scales in Table 2 with the CMS in Table 1 would reveal these differences.

4) How Can We Establish the Triple Implicational Scales?

The problem of how to establish an implicational relationship between the scale descriptions and the hierarchies of task difficulty is, in fact, the thorniest of the four problems to be settled in developing any CMS. It requires, in theory, *triple implicational scales*. One requirement is on an implicational scale between the scales of a CMS that a higher scale must subsume the lower ones, but not *vice versa*. Another is on an implicational scale between the scale descriptions that a higher description must subsume the lower ones, but not *vice versa*. The third requirement is on an implicational scale between the scale descriptions and the task hierarchies that the higher tasks or items mapping the higher descriptors must subsume the lower ones mapping the lower descriptors, but not *vice versa*.

In this respect, in the case of the Bachman-Palmer Scales on the IA approach as seen Table 2, the implicational scales between the scales, and between the descriptors, are regarded as having been empirically established, since both the scales and the descriptors have passed through a series of rather complicated calibration procedures as Bachman (1990) argued. Yet, though the scales and scale descriptors constitute an implicational relationship, it is still unclear how they can be translated directly into task specifications or items.

To follow the IL approaches would meet the triple implicational scales *in logic*. As seen

in Table 1, the PDS comprises the parallel developmental stages sequentially (i.e., 1X, 2X, ⋯, $1X^2$,⋯kX^3), which constitute an implicational scale as mentioned above. As we have already discussed, the IL approach enables us to map the PDS onto a CMS, and hence the absolute scales, the scale calibration, and the scale descriptions of the CMS can also be implicationally defined. The Testability hypothesis enables us to gear test tasks or items to each scale description of the CMS, and hence the resulting test tasks or items constitute an implicational hierarchy since the scale descriptions of the CMS are also in an implicational scale. Finally, the *Bounded-domain hypothesis* enables us to synchronize the test domains of the test tasks or items with the IL discourse domains on the continuum ranging from more concreteness to more abstractness. Thus, the scales, the scale descriptors, and the task or test hierarchies are all in an implicational relationship, and hence satisfy the *triple implicational scales* in logic.

In conclusion, not only can the IL approach provide a possible complement to the limitations inherent in the RL or IA approaches (i.e., indefinable absolute levels, imprecision or inconsistency of scale calibration and descriptions), but it also can satisfy what we call the triple implicational scales for developing a CMS. Thus, the IL approach enables us to convert the PDS into the CMS, and hence the CMS can function as an IL-sensitive or an acquisition-oriented scale for language pedagogy and SLA research.

III. RESEARCH METHOD

1. Participants

The participants of this study were 131 college students (18 males and 113 females), whose majors were linguistics (82%: Korean, English, French, Japanese, etc.), and the humanities and the social sciences (18%: history, philosophy, economics, etc.). The age of the participants ranged from 19 to 35, from freshmen to seniors.

2. Instruments[3]

The stimulus sentences were constructed according to the procedures adopted by Kim and Kwon (2005). From each Parallel Stage on the PDS (see Table 1), critical structures were first chosen. Based on these structures, five critical sentences were composed,

[3] This study may be a supplement to Kim's (2006). The study investigated the relationship of rating scales on the CMS to the scores on the GR and RC test, while this study investigates the relationship of those to the scores on the GR and LC test. For comparing the results of the two studies, this study used the same research method as that of Kim (2006).

which best reflected the parallel mechanisms of each stage of the PDS[4]. In order to minimize the effects of the length of the sentences and words on the experiment, each critical sentence was made up of "7±2" thought groups according to Miller's *magic number* (1956), using basic words and familiar proper nouns. These five critical sentences consisted of stimuli called the *critical sentence set* of forty-five items (see Appendix A).

The critical sentences were converted into two kinds of Sentence Construction Tests (SCT): the Sentence Conversion and the Sentence Combination Test. The SCT is a kind of IL-sensitive tests in that the items are geared to each stage of the PDS. These tests were played in *real time* onto the *Flash-window Method.* As shown in Figure 1 and 2, the method consisted of four phases (i.e., the Preparation, the Performance, the Completion, and the Relaxation Phase).

FIGURE 1

Sentence Conversion Test on the Flash-Window Method

Window 1: [Preparation]	[Preparing Instruction]: **Please read the following sentence carefully.** (A cursor will blink for 3 seconds, and then the following sentence will appear.) **John kissed Mary last night.** (This window will disappear in 2 seconds.)
Window 2: [Performance]	[Performing Instruction]: **Change '*last night*' into '*when*' in the following sentence, and make a *wh*-question in your mind.** (This window will disappear in 5 seconds, and then the following sentence will appear.) **John kissed Mary *last night*?** ↓ **When** (This window will disappear in 3 seconds.)
Window 3: [Completion]	[Completing Instruction]: **Write down the question you thought of on the answer sheet.** (This window will disappear in 15 seconds.)
Window 4: [Relaxation]	Along with a buzzer signaling the end of writing, music is played for about 10 seconds, and then another buzzer signals the start of the next item.

In the Preparation Phase, the participants were asked to read a given sentence or two or more sentences on the screen within a given time (160wpm). At the Performance Phase, in the case of the Sentence Conversion Tests, as seen in Figure 1, they were required to convert the sentence into another pattern (e.g., *wh*-questions) in their mind

[4] For example, at Parallel Stage 4X, the Parallel Mechanisms (i.e., Force-realizing, [-IFM] [-COM], Sentential [$_{CP}$...]) can produce such critical structures as *Wh*-inversion, *Neg*-inversion, and *Adv*-inversion as seen in Table 1. Of these structures, *Wh*-inversion is more prototypical and more widely used than the others. Thus, the *Wh*-inversion was chosen for constructing various types of critical sentences, that is, *Wh*-questions.

within a given time (120wpm). In the case of the Sentence Combination Tests, as in Figure 2, they were required to combine two or more sentences to form one complex sentence (e.g., indirect questions) in their mind within a given time. At the Completion Phase, they were asked to write down the converted or combined sentence they had just processed in their mind (60wpm). Finally, in the Relaxation Phase, music was played for about 10 seconds to help them to erase the previous item from their memories, to relieve their tension, and to prepare for the next item.

FIGURE 2

Sentence Combination Test on the Flash-Window Method

Window 1: [Preparation] [Preparing Instruction]: **Please read the following two sentences carefully.** (A cursor will blink for 3 seconds, and then the following sentences will appear.) **Do you know ...?** **Who kissed Mary last night?** (This window will disappear in 2 seconds.)
Window 2: [Performance] [Performing Instructions]: **Combine the next two sentences** **to make an indirect-question in your mind.** (This window will disappear in 5 seconds, and then the following will appear.) **Do you know + Who kissed Mary last night?** (This window will disappear in 4 seconds.)
Window 3: [Completion] [Completing Instructions]: **Write down the question you thought of on the answer sheet.** (This window will disappear in 20 seconds.)
Window 4: [Relaxation] Along with a buzzer signaling the end of writing, music is played for about 10 seconds, and then another buzzer signals the start of the next item.

Since the SCT was implemented in real time, the participants had to process such appropriate grammatical changes *in parallel* as constituent movements, word-order changes, and unifications of some grammatical features or morphemes to convert or combine sentences in real time. In this respect, this method can elicit more spontaneous responses from L2 learners than simple recall tests since it makes them rely less on their recall and more on actual processes in real time. Thus, the real time Flash-Window Method can provide a more sensitive measure for research into processability of IL development.

Also, in order to demonstrate the relationship of rating scales on the CMS and scores on a standardized test, this study adopted the Grammar (GR) and the Listening

Comprehension (LC) test of the TEPS. The LC test was used in an original form, but the GR test was slightly tailored for this research. Out of the 50 items on the test, the last 10 items were excluded because the types and the length of the items and the degree of difficulty were quite different from the other part of the test. As a result, the GR test was comprised of 40 multiple-choice items, and the LC test of 60 items.

3. Data Collection Procedures

The experiments were divided into two sections, each of which was carried out at a week interval. In the SCT section, participants were given some explanations of the SCT, and then presented a set of practice items through a beam-projector to familiarize them with the tests. After the practice items had been run, they were presented with the randomized test items. The test took about 60 minutes. After the test was completed, the participants were asked to fill out a demographic information questionnaire and an *excuse blank*: "I couldn't do my best in this test because of '_____' (such as having a cold, mental distraction, or reluctance)". The excuse blank was intended to give an opportunity to make any excuse for a participant who did not bias for best, and hence to exclude his or her data from the corpus. In the GR and LC test section, participants were told not to select an answer from the choices by "wild guessing". The GR test lasted 15 minutes and the LC test 60 minutes. After the two tests were finished, the participants were asked to fill out the questionnaires.

4. Data Analysis Procedures

From all the data elicited from the participants through the SCT, the followings were excluded in order to minimize the effects of faulty data on the result of the analysis: (a) irrelevant or unintelligible (illegible) cases, (b) cases unanswered to more than three items in a row, (c) cases filled out an *excuse blank*. The other data were analyzed by the following procedures. For the convenience of the analysis, to each response of each participant to each test item, a plus (+), a minus (-) or an equal (=) was given according to the degree of matching of a participant's response against the answer. A "+" was given to a response when it was matched to the answer. A "-" was given when it was either not matched to the answer or incomplete. If a response was not completely matched to the answer, and interim or transitional critical structures were *reconstructed* from it by the IL analysis, a "=" was given to the corresponding Parallel Stage which the reconstructed structures belong to on the PDS. Only trivial or insignificant lexical errors with no relations to the parallel mechanisms were overlooked. Criterion for *Pass* of each stage on the PDS was set as follows. If more than four responses of the five items of each stage match the

answers, a positive (+) was marked on the implicational scale, which means that the learner was considered passing the stage in question (called the *pass stage*). If less than four match the answers, a negative (-) was marked on the scale, which means that the learner was considered not passing the stage (called the *non-pass stage*)[5]. In the GR and LC test of the TEPS, one point was assigned to each item if it was a right answer, and then the number of right answers was tallied up.

IV. RESULTS AND DISCUSSION

1. The Correspondence of the Performance PDS to the Theoretical PDS

In order to see whether the theoretical Parallel Developmental Stages of the PDS reflect the actual developmental stages of the L2 learner, implicational scaling of the performance PDS was conducted. Of the responses elicited from the total 131 participants through SCT, four cases in the unanalyzable categories were excluded from the analysis. The others (127 cases) were analyzed by the Implicational Scaling according to the *Guttman* Procedures (Hatch & Lazaraton, 1991). The analysis showed that the coefficient of scalability (C_{scal}) was .82, far greater than the minimum requirement of .60 for an acceptable scale (see Appendix B). This scalability was almost the same as that (C_{scal}= .81) of Kim's study (2006), which reconfirmed that the performance PDS in real time by the L2 learners significantly corresponds to the theoretical PDS.

2. The Relationship of the Rating Scales on the CMS to the Scores on Proficiency Test

In order to see if there is a relationship between the rating scales on the CMS and the scores on the GR and LC test of the TEPS, the Spearman Correlation Test and the Kruskal-Wallis Test were conducted (see the last two columns in Appendix B). The results of Spearman Correlation are illustrated in Table 3. The ratings on the CMS were significantly correlated with the scores on the GR test ($r=.83$, $p<.01$), but moderately with the scores on the LC test ($r=.61$, $p<.01$). Referring to the result of Kim (2006), the ratings on the CMS were also highly correlated with the scores on the GR test ($r=.90$, $p<.01$), but fairly with the scores on the RC test ($r=.69$, $p<.01$).

[5] This research simply followed the convention that requires five instances and an 80% cut off points, though there is no well-documented rationale for either of these conventions (Hatch & Lazaraton, 1991, p. 215).

TABLE 3
Correlations between CMS and GR/LC

	CMS	GR	LC
CMS	1	.83*	.61
GR		1	.66
LC			1

Note: *p<.01

According to the results of the two studies (this study & Kim (2006)), the coefficients of the LC and the RC test were much lower than that of the GR test. One reason may result from the fact that LC and RC tests are generally involved in much more variable test domains than GR tests. In other words, reading and listening materials encompass much more divergent contexts, contents, topics, texts, and genres than GR tests. Another may come from the fact that LC and RC tests are in general related to much more various dimensions of language use and skills such as prior knowledge, complex cognitive processes, reading and listening skills and strategies than GR tests. The coefficient of the CMS to the LC was the lowest in the three proficiencies, which may result from the fact that, unlike GR and RC tests, LC tests generally require the understandings of spoken as well as written language. Even if the coefficients ranged from .61 to .90 depending on the proficiency tests, the analysis of the Correlation Test showed significant relationships of the CMS with language proficiencies of the GR, the LC, and the RC test. This suggests that the CMS has a potential to be used as an IL-sensitive scale for language teaching, language testing, and SLA research.

TABLE 4
Kruskal-Wallis Test Statistics (a, b)

	GR	LC
Chi-Square	64.18	48.58
Df	9	9
Asymp. Sig	.000	.000

Note: a. Kruskal-Wallis Test, b. Grouping Variable: CMS

A more in-depth analysis, the Kruskal-Wallis Test was conducted to examine whether there is a rank difference among levels of the CMS. As shown in Table 4, the value for Kruskal-Wallis χ^2 stood at 64.18 for the GR and 48.58 for the LC. In addition, as shown in Table 5, the order of levels of the CMS came into line with the order of mean ranks of the GR and the LC test. According to Kim's study (2006), the value for Kruskal-Wallis χ^2 was at 83.17 for the GR and 50.10 for the RC, and the levels of the CMS was also squarely held in the order of mean ranks of the RC. The analysis of the Kruskal-Wallis Test proved a

statistical difference among the levels of the CMS. In conclusion, this study corroborated the finding of Kim's research (2006) that the CMS based on the PDS has its own validity.

TABLE 5
Kruskal-Wallis Test: Ranks

PDS Stages	CMS Levels	N (93)	Mean Rank GR	N (120)	Mean Rank LC
1X	0	1	1.50	1	1.00
2X	1	4	8.88	4	13.88
3X	2	9	18.56	11	25.91
4X	3	10	25.20	12	41.71
$1X^2$	4	16	34.47	22	54.80
$2X^2$	5	14	43.54	18	58.00
$3X^2$	6	12	62.04	16	70.47
$4X^2$	7	15	70.03	19	73.66
$5X^2$	8	10	77.65	14	92.64
KX^3	9	2	91.25	3	114.83

V. CONCLUSION AND IMPLICATIONS

On the Parallel Developmental Sequence (PDS) Model (Kim & Kwon, 2005, 2007), Kim (2006) proposed the Common Metric Scale (CMS) Model for measuring language ability. The CMS was developed in terms of the IL approach, the Testability and the Bounded-domain hypothesis. The IL approach enables us to map the PDS onto the CMS, and hence the absolute scales, the scale calibration, and the scale descriptions of the CMS can be automatically defined. The Testability hypothesis enables us to gear test tasks to the Parallel Stages of the PDS along which IL learners proceed from the initial stage through the next higher stage, in a stepwise fashion, until up to the final stage. The Bounded-domain hypothesis enables us to synchronize test domains with IL discourse domains on the continuum ranging from more concreteness to more abstractness along which IL learners create and expand in their IL use. Thus, the scales, the scale descriptors, and the task or test hierarchies are all in an implicational relationship, and hence satisfy what we call the *triple implicational scales*. To validate the theoretical CMS, Spearman Correlation and the Kruskal-Wallis Test were conducted. The analyses of the tests demonstrated that the CMS holds statistical significance and has its validity. In conclusion, not only can the CMS on the PDS constitute a parallel interface to enhance the reciprocal developments between SLA research and language testing, but it also can function as an *IL-sensitive* or an *acquisition-oriented scale* for identifying IL learners' current developmental stages, diagnosing the deficiencies in their IL grammar, and predicting what rules or structures will

be learnable or teachable next, thereby giving them a "timely remedy and tailored teaching".

Given these possibilities, how can we implement the CMS into language teaching and testing in practice, thereby providing a "timely remedy and tailored teaching" for IL learners? As Kim (2007) suggested, there are at least three ways for implementing the CMS into language testing depending on the kinds of language tests such as IL-sensitive tests, task/function-based tests, and comprehension tests.

IL-sensitive tests, as the name suggests, are designed to be sensitive to or tailored for the Parallel Stages on the PDS in terms of the IL approach. These tests include Sentence Construction Tests, Recall Speaking Tests, IL-sensitive cloze tests, and others (Kim, 2007). The implementation of the CMS into IL-sensitive tests first involves analyzing IL learners' performance elicited from administrating the IL-sensitive tests, and determining what stage the IL structures produced by the learners belong to on the PDS. This is followed by an assessment of their IL proficiency and a diagnosis of their deficiency in terms of the CMS. As already shown in the real time SCT on Flash-Window Method, IL-sensitive tests are so sensitive to each stage of the PDS that they help us to make more accurate diagnoses and more timely remedies. Yet, they are structured so narrowly that they are less likely to capture interaction, and hence to produce authentic and interactional data. Thus, IL-sensitive tests need to be complemented by task/function-based tasks.

Task/function-based tasks are designed to elicit or collect more authentic IL data through interactions of IL learners with each other or a researcher/tester. These tasks include various kinds of task/function-based tasks, a set of writing tasks, oral interviews, and others (Kim, 2007). In task/function-based tasks, the CMS is implemented drawing on the *Profile Analysis Procedure* (Clahsen, 1985), or, more recently, the *Rapid Profile* (Pienemann & Mackey, 1993). As Pienemann (1998) illustrated in his task variation research, the procedure first involves collecting samples of IL production through the administration of task/function-based tasks creating two-way interactive communications. This is followed by an analysis of IL learners' output using the procedure which matches the critical structures and the error patterns produced against the parallel stages on the PDS, and by an assessment of their IL proficiency in terms of the CMS.

In comprehension tests (listening/reading), first, in accordance with the Testability hypothesis, the passages for listening or reading tests are selected, or, if necessary, tailored, thereby the grammatical structures of each passage being geared to the critical structures of each stage of the PDS. Then, in accordance with the Bounded-domain hypothesis, as discussed earlier, the test domains of each passage are synchronized with the IL discourse domains, thereby test domains of each passage being interlocked with the IL discourse domains on the continuum ranging from more concrete domains to more abstract ones. Finally, the comprehension tests are administered, and then the scores of the tests are converted into and

reported on a rating on the CMS.

On the basis of the assessment, the diagnoses, and the remedies resulting from the implementation of the CMS into various tests, we can provide IL learners with learning tasks that are tailored for their developmental stages and focused on their deficiencies, with employing teaching methods that are well suited to the characteristics of learning tasks to be presented, their learning styles, and teaching situations, etc. In conslusion, this is a way we can implement the CMS into language teaching and testing in practice, and hence we can provide IL learners with a "timely remedy and tailored teaching". Yet, this invites us to do more research into a more principled and viable way for establishing implicational hierarchies between the levels of the CMS and the hierarchies of test contexts, contents, and tasks, thus maximizing the "fit" between the levels of test takers' language ability and the hierarchies of item or task difficulty.

REFERENCES

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

Bachman, L. F., & Cohen, A. D. (1998). Language testing-SLA research interface: An update. In L.F. Bachman & A. D. Cohen (Eds.), *Interface between second language acquisition and language testing research* (pp. 1-31). Cambridge: Cambridge University Press.

Bachman, L. F., & Palmer, A. S. (1983). *Oral interview test of communicative proficiency in English.* Los Angeles, CA: Photo-offset.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests.* Oxford: Oxford University Press.

Brindley, G. P. (1986). *The assessment of second language proficiency: Issues and approaches.* NSCR Research Series. Adelaide: National Curriculum Resource Center.

Brindley, G. P. (1998). Describing language development? In L.F. Bachman & A. D. Cohen (Eds.), *Interface between second language acquisition and language testing research* (pp. 112-140). Cambridge: Cambridge University Press.

Clahsen, H. (1985). Profiling second language development: A procedure for assessing L2 proficiency. In K. Hyltenstam & M. Pienemann (Eds.), *Modeling and assessing second language acquisition* (pp. 283-325). Clevedon: Multilingual Matters.

Clark, J. L. D. (1980). Toward a common measurement of speaking proficiency. In J. R. Firth (Ed.), *Measuring spoken language proficiency* (pp. 15-26). Washington, DC: Georgetown University Press.

Douglas, D., & Selinker, L. (1985). Principles for language tests within the "discourse domains" theory of interlanguage: Research, test construction and interpretation. *Language Testing, 2*, 205-226.

Douglas, D. (1998). Testing methods in context-based SL research. In L.F. Bachman & A. D. Cohen (Eds.), *Interface between second language acquisition and language testing research* (pp. 141-155). Cambridge: Cambridge University Press.

Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics.* New York: Newbury House.

Ingram, D. (1985). Assessing proficiency: An overview on some aspects of testing. In K. Hyltenstam & M. Pienemann (Eds.), *Modeling and assessing second language acquisition* (pp. 215-276). Clevedon: Multilingual Matters.

Kim, Boram. (2006). The acquisition of the English locative alternation by Korean EFL learners: What makes L2 learning difficult? *English Language & Literature Teaching, 12*(4), 31-68.

Kim, Y. M. (2006). A common metric scale (CMS) on the parallel developmental sequence (PDS) model. *English Teaching, 61*(4), 77-107.

Kim, Y. M. (2007). Diagnosis and remedy systems (DRS) for teaching English based on the common metric scale (CMS) model. *English Teaching, 62*(2), 47-77.

Kim, Y. M., & Kwon, O. (2005). *A parallel developmental sequence.* Paper presented at the 5th International Symposium on Processability, Second Language Acquisition and Bilingualism. Melbourne, Australia.

Kim, Y. M., & Kwon, O. (2007). A parallel developmental sequence in second language acquisition. In F. Mansouri (Ed.), *Second language acquisition research: Theory-construction and testing* (pp. 239-275). Newcastle: Cambridge Scholars Press.

Long, M, H., & Richards, J. C. (1998). Series editor's preface. In L. F. Bachman & A. D. Cohen (Eds.), *Interface between second language acquisition and language testing research* (pp. 112-140). Cambridge: Cambridge University Press.

Meisel, J., Clahsen, H., & Pienemann, M. (1981). On determining developmental stages in natural second language acquisition. *Studies in Second Language Acquisition, 3*(2), 109-135.

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity of processing information. Psychological Review, 63, 81-97.

Park, Hyeson. (2006). Acquisition of English complex predicates in SLA. *English Language & Literature Teaching, 12*(2), 177-194.

Pienemann, M. (1985). Psychological constraints on the teachability of languages. *Studies in Second Language Acquisition, 6,* 186-214.

Pienemann, M. (1998). *Language processing and second language development: Processability*

theory. Amsterdam: John Benjamins.

Pienemann, M., & Johnston, M. (1987). Factors influencing the developmental of language proficiency. In D. Nunan (Ed.), *Applying second language acquisition research* (pp. 45-141). Adelaide, Australia: National Curriculum Resource Center.

Pienemann, M., & Mackey, A. (1993). *An empirical study of children's ESL development and Rapid Profile.* Sydney: NLLIA Language Acquisition Research Center, University of Sydney.

Shohamy, E. (1998). Language testing and SLA. In L.F. Bachman & A. D. Cohen (Eds.), *Interface between second language acquisition and language testing research* (pp. 156-176). Cambridge: Cambridge University Press.

Vainikka, A., & Young-Scholten, M. (1998). The initial state in the L2 acquisition of phrase structure. In S. Flynn, G. Martohardjono & W. O'Neil, (Eds.), *The generative study of second language acquisition* (pp. 17-34). Mawha, NJ: Lawrence Erlbaum.

Appendix A

The Critical Sentence Set

1. Parallel Developmental Stage 1X
2. Parallel Developmental Stage 2X
   1) Do you wash the dishes everyday?
   2) Who opened the windows on this cold day?
   3) What happened at the party last night?
   4) Who appeared on the concert last night?
   5) Do you play tennis with your friends every weekend?
3. Parallel Developmental Stage 3X
   1) Did you watch that news on TV last night?
   2) Can you go camping with us this weekend?
   3) Have you read an English novel recently?
   4) Do you want to go there this weekend?
   5) Would you like to have a snack after studying?
4. Parallel Developmental Stage 4X
   1) When did you see her on the campus?
   2) Where would you like to have a drink after the show?
   3) What have you been writing since last week?
   4) Who do you want to meet at the concert this Friday?
   5) What did you give to Mary at the party last night?

5. Parallel Developmental Stage X^2

5.1. Parallel Developmental Stage $1X^2$

1) Did you ask her to go for a drive last night?
2) What does he want you to do this summer vacation?
3) Who does she expect to hold a party next time?
4) Who does he want her to meet this Sunday?
5) I think that John is going with his sweetheart to the concert today.

5.2. Parallel Developmental Stage $2X^2$

1) Did you ask your daughter who liked her in her class?
2) The man who studies mechanics at college will investigate the car accident.
3) Do you know who taught English to this class last year?
4) Have you ever met the lady who was talking with John in the park last night?
5) Do you know who helped him to fix the broken TV?

5.3. Parallel Developmental Stage $3X^2$

1) John has proposed to the girl whom you longed to meet at college.
2) The girl whom you taught in college is going to marry my son next month.
3) The man whom Mary spoke to in the theater studies history at the college.
4) Have you ever met the man whom my father expects me to marry?
5) John employed a woman whom he thought to be honest at the job interview.

5.4. Parallel Developmental Stage $4X^2$

1) Does she know whom John came with to the party last night?
2) Do you want to know why Jane decided to divorce her husband last year?
3) Would you like to know who my mother wants me to meet this Sunday?
4) Do you want to know who Tom ordered to fire her after the audit?
5) Did you ask Mary what she handed John during history class?

5.5. Parallel Developmental Stage $5X^2$

1) Why do you think John parted from his sweetheart last week?
2) Where do you think they went to have drinks after the show?
3) Who does she think John handed the note to during the English test?
4) Who do you think Mary expects him to invite to the reception?
5) Who do you suppose John wants to introduce her to the mayor?

6. Parallel Developmental Stage X^3

1) Do you want to know who John thinks Mary gave the nice gift to at the party?
2) Who do you guess Mary thinks John loves at the college?
3) Do you know who Mary thinks John asked Alice to invite to the reception?
4) Who do you believe John thinks Mary wants to introduce Alice to the president?
5) Who do you think John believes Mary told Alice to fire after the audit?

Appendix B

The Implicational Scaling Scalogram

Subjects	Parallel Developmental Stages										CMS	TEPS
	kX^3	$5X^2$	$4X^2$	$3X^2$	$2X^2$	$1X^2$	4X	3X	2X	1X	Level	GR/LC
74	+	+	+	+	+	+	+	+	+	+	9	33/52
116	+	+	+	+	+	+	+	+	+	+	9	x/56
54	+	+	+	+	+	+	+	+	+	+	9	34/48
18	-	+	+	+	+	+	+	+	+	+	8	28/39
31	-	+	+	+	+	+	+	+	+	+	8	25/41
34	-	+	+	+	+	+	+	+	+	+	8	31/47
36	-	+	+	+	+	+	+	+	+	+	8	28/42
37	-	+	+	+	+	+	+	+	+	+	8	34/51
38	-	+	+	+	+	+	+	+	+	+	8	30/46
49	-	+	+	+	+	+	+	+	+	+	8	21/43
67	-	+	+	+	+	+	+	+	+	+	8	x/39
68	-	+	+	+	+	+	+	+	+	+	8	26/48
88	-	+	+	+	+	+	+	+	+	+	8	33/49
96	-	+	+	+	+	+	+	+	+	+	8	x/44
98	-	+	+	+	+	+	+	+	+	+	8	x/48
123	-	+	+	+	+	+	+	+	+	+	8	x/44
125	-	+	+	+	+	+	+	+	+	+	8	33/48
01	-	-	+	+	+	+	+	+	+	+	7	24/53
07	-	-	+	+	+	+	+	+	+	+	7	x/33
09	-	-	+	+	+	+	+	+	+	+	7	x/47
10	-	-	+	+	+	+	+	+	+	+	7	22/32
40	-	-	+	+	+	+	+	+	+	+	7	28/28
44	-	-	+	+	+	+	+	+	+	+	7	x/41
45	-	-	+	+	+	+	+	+	+	+	7	23/35
55	-	(+)	+	(-)	+	+	+	+	+	+	7	30/45
56	-	-	+	+	+	+	+	+	+	+	7	25/45
57	-	-	+	+	+	+	+	+	+	+	7	31/39
60	-	-	+	+	+	+	+	+	+	+	7	28/x
63	-	-	+	+	+	+	+	+	+	+	7	x/38
77	-	-	+	+	+	+	+	+	+	+	7	28/37
78	-	-	+	+	+	+	+	+	+	+	7	32/48
79	-	-	+	+	+	+	+	+	+	+	7	29/49
94	-	-	+	+	+	+	+	+	+	+	7	x/x
102	-	-	+	+	+	+	+	+	+	+	7	25/49
106	-	-	+	+	+	+	+	+	+	+	7	26/36
110	-	-	+	+	+	+	+	+	+	+	7	22/39
112	-	-	+	+	+	+	+	+	+	+	7	26/46
114	-	-	+	+	+	+	+	+	+	+	7	x/37
121	-	-	+	+	+	+	+	+	+	+	7	x/x
02	-	-	(+)	(-)	+	+	+	+	+	+	6	24/46
13	-	-	(+)	(-)	+	+	+	+	+	+	6	25/42
26	-	-	(+)	(-)	+	+	+	+	+	+	6	x/25
32	-	-	+	+	+	+	+	+	+	+	6	30/40
39	-	-	+	+	+	+	+	+	+	+	6	23/36
42	-	(+)	-	(-)	+	+	+	+	+	+	6	x/46
48	-	-	(+)	(-)	+	+	+	+	+	+	6	x/x
53	-	-	(+)	(-)	+	+	+	+	+	+	6	x/36
58	-	-	(+)	(-)	+	+	+	+	+	+	6	25/46
65	-	-	-	+	+	+	+	+	+	+	6	27/40
71	-	-	(+)	(-)	+	+	+	+	+	+	6	27/51
90	-	-	-	+	+	+	+	+	+	+	6	28/42
100	-	-	-	+	+	+	+	+	+	+	6	22/33

101	-	-	-	+	+	+	+	+	+	+	6	22/39
105	-	-	-	+	+	+	+	+	+	+	6	24/46
108	-	-	-	+	+	+	+	+	+	+	6	22/30
109	-	-	-	+	+	+	+	+	+	+	6	x/41
08	-	-	-	-	+	+	+	+	+	+	5	21/43
16	-	-	-	-	+	+	+	+	+	+	5	22/36
23	-	-	-	-	+	+	+	+	+	+	5	24/44
27	-	-	-	-	+	+	+	+	+	+	5	21/52
50	-	-	-	(+)	+	(-)	+	+	+	+	5	17/32
61	-	-	-	-	+	+	+	+	+	+	5	20/29
62	-	-	-	-	+	+	+	+	+	+	5	16/23
64	-	-	-	-	+	+	+	+	+	+	5	x/45
72	-	-	-	-	+	+	+	+	+	+	5	25/40
84	-	-	(+)	-	(-)	+	+	+	+	+	5	21/33
86	-	-	(+)	-	(-)	+	+	+	+	+	5	21/38
87	-	-	(+)	-	(-)	+	+	+	+	+	5	24/30
92	-	-	-	-	+	+	+	+	+	+	5	24/37
95	-	-	(+)	-	(-)	+	+	+	+	+	5	x/32
104	-	-	-	-	+	+	+	+	+	+	5	23/44
115	-	-	-	-	+	+	+	+	+	+	5	x/48
119	-	-	-	-	+	+	+	+	+	+	5	24/37
122	-	-	(+)	-	(-)	+	+	+	+	+	5	x/32
06	-	-	-	-	(+)	(-)	+	+	+	+	4	19/43
15	-	-	-	-	-	+	+	+	+	+	4	19/38
17	-	-	-	-	-	+	+	+	+	+	4	16/29
20	-	-	-	-	-	+	+	+	+	+	4	18/42
22	-	-	-	-	-	+	+	+	+	+	4	x/34
28	-	-	-	-	-	+	+	+	+	+	4	20/33
29	-	-	-	-	-	+	+	+	+	+	4	24/50
30	-	-	-	-	-	+	+	+	+	+	4	25/29
21	-	-	-	-	(+)	(-)	+	+	+	+	4	14/34
59	-	-	-	-	-	+	+	+	+	+	4	19/45
66	-	-	-	-	-	+	+	+	+	+	4	17/31
73	-	-	-	-	-	+	+	+	+	+	4	x/32
75	-	-	-	-	-	+	+	+	+	+	4	18/32
83	-	-	-	(+)	-	(-)	+	+	+	+	4	18/33
85	-	-	-	-	-	+	+	+	+	+	4	20/45
93	-	-	-	-	-	+	+	+	+	+	4	x/38
97	-	-	-	-	-	+	+	+	+	+	4	25/33
113	-	-	-	-	(+)	(-)	+	+	+	+	4	x/41
117	-	-	-	-	-	+	+	+	+	+	4	22/43
118	-	-	(+)	(+)	-	(-)	(-)	+	+	+	4	x/35
120	-	-	-	-	-	+	+	+	+	+	4	x/38
124	-	-	-	-	-	+	+	+	+	+	4	25/33
04	-	-	-	-	-	-	+	+	+	+	3	18/39
05	-	-	-	-	-	-	+	+	+	+	3	x/44
11	-	-	-	-	-	-	+	+	+	+	3	18/33
12	-	-	-	-	-	-	+	+	+	+	3	16/20
24	-	-	-	-	-	-	+	+	+	+	3	18/33
33	-	-	-	-	-	-	+	+	+	+	3	x/40
43	-	-	-	-	-	(+)	(-)	+	+	+	3	23/35
52	-	-	-	-	-	-	+	+	+	+	3	20/30
41	-	-	-	-	-	-	+	+	+	+	3	14/20
70	-	-	-	-	-	-	+	+	+	+	3	23/48
81	-	-	-	-	-	-	+	+	+	+	3	x/x
107	-	-	-	-	-	-	+	+	+	+	3	13/21
111	-	-	-	-	-	-	+	+	+	+	3	19/26
03	-	-	-	-	-	-	-	+	+	+	2	15/24
14	-	-	-	-	-	-	-	+	+	+	2	x/25

19	-	-	-	-	-	-	-	+	+	+	2	20/24
35	-	-	-	-	-	-	-	+	+	+	2	21/35
46	-	-	-	-	-	-	-	+	+	+	2	15/33
51	-	-	-	-	-	-	-	+	+	+	2	20/28
69	-	-	-	-	-	-	-	+	+	+	2	13/36
76	-	-	-	-	-	-	-	+	+	+	2	18/35
80	-	-	-	-	-	-	-	+	+	+	2	15/33
99	-	-	-	-	-	-	-	+	+	+	2	x/29
103	-	-	-	-	-	-	-	+	+	+	2	14/20
126	-	-	-	-	-	-	-	+	+	+	2	14/x
127	-	-	-	-	-	-	-	+	+	+	2	12/x
25	-	-	-	-	-	-	-	-	+	+	1	x/26
47	-	-	-	-	-	-	-	-	+	+	1	17/23
82	-	-	-	-	-	-	-	-	+	+	1	x/25
89	-	-	-	-	-	-	-	-	+	+	1	15/31
91	-	-	-	-	-	-	-	-	-	+	0	12/19
Total	124	3 108	19 73	54 77	50 55	72 36	91 20	107 5	122 1	126 0	127	

+: Pass Stage, -: Non-pass Stage, (+)/(-): Non-parallel Stage, x: Non-attendance at TEPS

Examples in: English
Applicable Languages: English
Applicable Levels: all learners from children to adults

Yong Myeong Kim
Hankuk University of Foreign Studies
201-ho, Dukmoon-building, 106-1, Dongbinggo-dong, Yongsan-gu
Seoul 140-811, Korea
Email: oprcms@yahoo.co.kr

Received in Jan., 2008
Reviewed in Feb., 2008
Revised version received in Mar., 2008