

Journal of the National Council of Less Commonly Taught Languages

Volume 27



National Council of Less Commonly Taught Languages

Danko Šipka, Editor
Antonia Schleicher, Managing Editor
Charles Schleicher, Copy Editor
Olutayo Ajayi, Production Editor

Please address inquiries concerning advertising, subscription, and
issues to the NCOLCTL Secretariat at the following address:

National Council of Less Commonly Taught Languages
601 S Woodscrest Dr. Bloomington, IN 47401

Copyright © 2020, National Council of Less Commonly Taught Languages
(NCOLCTL)

The Journal of the National Council of Less Commonly Taught Languages, published annually by the Council, is dedicated to the issues and concerns related to the teaching and learning of Less Commonly Taught Languages. The Journal primarily seeks to address the interests of language teachers, administrators, and researchers. Articles that describe innovative and successful teaching methods that are relevant to the concerns or problems of the profession, or that report educational research or experimentation in Less Commonly Taught Languages are welcome. Papers presented at the Council's annual conference will be considered for publication, but additional manuscripts from members of the profession are also welcome.

Besides the Journal Editor, the process of selecting material for publication is overseen by the Advisory Editorial Board, which consists of the foremost scholars, advocates, and practitioners of LCTL pedagogy. The members of the Board represent diverse linguistic and geographical categories, as well as the academic, government, and business sectors.

JNCOLCTL Advisory Editorial Board

Elizabeth M. Bergman

Miami University of Ohio, Arabic

Richard Brecht

*American Council for International Education, Russian and Language
Advocacy*

Stephane Charitos

*Columbia University, Language Technology,
Greek*

Victor Friedman

*University of Chicago, Balkans and Caucasus
Languages*

Michael Hillmann

University of Texas-Austin, Farsi, Tajik

Scott McGinnis

*Defense Language Institute, Washington DC,
Chinese*

Rajandra Singh

University de Montreal, South Asian languages, Theoretical Approaches

The Journal of the National Council of Less Commonly Taught Languages is a refereed publication. The editors invite submissions of articles, book reviews, and advertisements for future volumes. To be considered for publication, manuscripts should be typed, double spaced, with margins of at least one and one-half inches. Articles should include a 150-word abstract and should be about 20 typewritten pages. All references and notes should follow the latest edition of the Publication Manual of the American Psychological Association (APA)
See <http://www.apastyle.org/manual/index.aspx>

Manuscripts should be:

- A maximum of 20 pages (excluding references, charts, notes, etc.) and submitted electronically as a Word-document via e-mail attachment.
- Double-line spaced throughout, including notes, references, and tables, using 12-point **Garamond** font with an inch left margin. (*Please ensure that this specified formatting is followed*).
- Accompanied by a 150 word (or less) abstract and a cover sheet containing the manuscript title, name, address, office and home telephone numbers, fax number, email address, and full names and institutions of each author. (*Because the manuscript will be blind reviewed, identifying information should be on the cover sheet only, and not appear in the manuscript*).

All manuscripts should be electronically submitted via email to the Editor at: Ebru.Turker@asu.edu and copied to the Secretariat at: secretariat@ncolctl.org.

Subscriptions and address changes to the journal should be sent to

Dr. Antonia Schleicher
National Council of Less Commonly Taught Languages
601 S Woodcrest Dr, Bloomington, IN 47401.

Subscription Rates:

1 Year	Individual	U.S. \$60
1 Year	Institutional	U.S. \$200

**Journal of the National
Council of Less Commonly
Taught Languages**

Volume 27

Table of Contents

<i>Editor's Introduction</i> Danko Sipka	ix
Dmitrii Pastushenkov, Tanya McIntyre <i>Life after Language Immersion: Two Very Different Stories</i>	1
Donglin Chai, Bing Mu <i>Connecting Language Learning in the Classroom with the Local Community: Using Field Performance Tasks in Chinese Study Abroad Contexts</i>	33
Cheng-Fu Chen <i>Temporal Sequencing and Narration in Learner Language: The Case of an Intensive Mandarin Chinese Program</i>	74
Hye-Sook Wang <i>An Analysis of Testing Practices in College Korean Language Classrooms</i>	130
Margherita Berti <i>Cultural Representations in Foreign Language Textbooks: A Need for Change</i>	175

Authors' Contact Information

1. Dmitrii Pastushenkov, Tanya McIntyre *Life after Language Immersion: Two Very Different Stories* -
pastushe@msu.edu; tbystro2@kent.edu
2. Donglin Chai, Bing Mu *Connecting Language Learning in the Classroom with the Local Community: Using Field Performance Tasks in Chinese Study Abroad Contexts* -
chai.39@buckeyemail.osu.edu
3. Cheng-Fu Chen *Temporal Sequencing and Narration in Learner Language: The Case of an Intensive Mandarin Chinese Program* -
cchen5@olemiss.edu
4. Hye-Sook Wang *An Analysis of Testing Practices in College Korean Language Classrooms* -
hye-sook_wang@brown.edu
5. Margherita Berti *Cultural Representations in Foreign Language Textbooks: A Need for Change* -
berti@email.arizona.edu

National Council of Less Commonly Taught Languages (NCOLCTL)

NCOLCTL is an organization dedicated to the teaching and learning of Less Commonly Taught Languages. Membership is open to individuals and organizations that share this interest.

NCOLCTL Homepage

<http://www.ncolctl.org>

NCOLCTL Officers

President: An Chung Cheng, *University of Toledo*

Vice-President: Luis Gonçalves, *Princeton University*

Secretary/Treasurer: Susan Schmidt, *University of Colorado*

Executive Director: Antonia Schleicher, *Indiana University*

Executive Board:

Patrica Sabarwal - Member-at-Large
University of Kansas

Hye Young Shin – Member-at-Large
American University

David Bong - Member-at-Large
CEO Avant Assessment

NCOLCTL Mission Statement

NCOLCTL's mission is to increase the number of Americans who choose to learn one or more of the less commonly taught languages (LCTLs) as a means of enhancing cross-cultural communication among citizens of the United States. NCOLCTL's work focuses on the less commonly taught languages which are becoming increasingly vital to the economic, social and political welfare of the United States. Furthermore, NCOLCTL seeks to improve the teaching and learning of these languages and to make them more generally available. We are the national voice for organizations and individuals who represent the teaching of these less commonly taught languages.

Both the collegiate and precollegiate level. NCOLCTL also promotes the use of technology, especially computers and the Internet, to enable a new era in cross-cultural understanding, communication, and language education.

NCOLCTL constitutes a national mechanism devoted to strengthening the less commonly taught language professions through enabling NCOLCTL members to work toward “shared solutions to common problems.” NCOLCTL principally directs its efforts toward building a national architecture for the LCTL field and in making the field’s resources easily accessible to language programs and individual learners around the United States.

Activities of NCOLCTL

Since its establishment in 1990, the NCOLCTL has carried out a variety of activities to raise awareness about the importance of less commonly taught languages. NCOLCTL achieves its goals through the following activities:

- Annual Conference in conjunction with Delegate Assembly
- Conducting research to promote and facilitate the learning and teaching of the LCTLs
- Planning for and establishing a national policy for building the national capacity for the study of the LCTLs
- Enhancing the capacity of existing LCTL national associations, and organizing new ones
- Establishing a system for networking and communication among member organizations, and facilitating their collective efforts to solve problems in the LCTL field
- Developing language learning frameworks to guide teacher training, curriculum design, materials development, and seek ways to address problems of articulation among different levels of the American educational system
- Working, on behalf of the members, with government agencies, foundations, and the general foreign language community on policy issues and to seek funding to establish effective standards for the less commonly taught language field
- Fostering national and international linkages within and across the various language areas

- Online Teaching Courses designed primarily for new instructors of LCTLs at postsecondary level and a useful resource for experienced instructors.
-

Becoming a member of NCOLCTL

Both individual and organizational Membership Application forms for the National Council of Less Commonly Taught Languages are available online at:

<http://www.ncolctl.org/membership>

Editor's Introduction

Danko Šipka
Arizona State University

The spring 2020 issue features eight papers and one review article, representing various topics of interest to the entire NCOLCTL community and various languages in the field, and it comes in two volumes. In this volume, the first two papers *Life after Language Immersion: Two Very Different Stories*, and *Connecting Language Learning in the Classroom with the Local Community: Using Field Performance Tasks in Chinese Study Abroad Contexts* discuss immersion and study abroad, programs that augment our regular classroom activities. The next paper, titled *Temporal Sequencing and Narration in Learner Language: The Case of an Intensive Mandarin Chinese Program*, discusses the issue of narration, one of the key problems in presentative speaking and writing. Testing practices are front and center in *An Analysis of Testing Practices in College Korean Language Classrooms*. This volume concludes with a particularly interesting review article, entitled *Cultural Representations in Foreign Language Textbooks: A Call for Change*.

An Analysis of Testing Practices in College Korean Language Classrooms¹

Hye-Sook Wang

Brown University

Abstract

The aim of this study is to examine current testing practices in Korean language classrooms in U.S. colleges and universities. Twelve final examinations of beginning and intermediate level classes are analyzed following Bachman & Palmer's 'task characteristics' as an analytical tool. The results show that first, the majority of tests includes listening, reading, and writing, while a few tests focus exclusively on grammar and vocabulary. Second, the relative proportion of the four skills varies notably from school to school and varies less between levels. Third, the sequence of parts/tasks is very similar in all tests regardless of program or level. Fourth, selected response and limited production are the most popular response types in the grammar/vocabulary section of all the samples. There are virtually no extended response questions. Fifth, the length of input in reading and listening passages varies notably from program to program, although the results are hard to generalize for level differences. Implications of these findings are discussed.

Keywords: Teaching Korean as a Foreign Language; Assessment and testing; Korean language classrooms; Achievement test; Final examination

¹ The original version of this paper was presented at the annual conference of the American Association of Teachers of Korean at Boston University in June 2014.

Introduction

Assessment is one of the most important aspects and crucial activities that any teachers engage in, be it through tests or other means. Teachers spend a significant amount of time both in developing the assessment tools (e.g., tests, rubrics) as well as grading/evaluating them. As such, a number of fundamental questions need to be asked: a) how do teachers make tests as clear and unambiguous as possible, and truly ‘useful’ and ‘meaningful’ to serve the intended purposes; b) how do teachers measure success in terms of the relationship between teaching (i.e., what they instruct) and testing (i.e., what they evaluate/assess), especially in classroom-based achievement tests; c) how can challenges in test development be effectively addressed and overcome from a practical perspective; and d) how do teachers assess students’ actual learning as well as their knowledge or skill of testing—two separate issues.

The relative importance of tests in a given course may depend on various factors, such as the level and objectives of the course. But a general practice for many language teachers as described in their course syllabi appears to be that tests (e.g., unit/lesson tests, mid-term exams, final exams) are given relatively heavy weight, if not heaviest, compared to other course requirements such as attendance and participation, projects, homework, or quizzes, etc. This is especially true in the case of lower-level courses. Naturally, students seem to be gravely concerned about and sensitive to the impact of test scores on their overall course grade in addition to or irrespective of their actual learning outcomes. In this respect, fair and effective test administration on the part of the teachers as test developers—the entire process from design, development, to grading a test—becomes even more critical. Despite the significance of this topic in and of itself, little attention has been paid to this aspect in the field of Korean as

a Foreign Language (KFL henceforth), evidenced by a small volume of publications (see the literature review in the next section). While test development, administration, and grading are seen as a routine practice for teachers, perhaps one of the reasons any testing-related activities are taken for granted, it is time to critically reflect on current practices.

We analyze tests for a number of reasons and purposes. First, testing is one of the most essential and crucial components of teaching and learning processes in a classroom setting, and arguably the most laborious and challenging task for teachers given its consequences. Unlike other performance-based activities and tasks that learners produce (e.g., essays, projects, presentations, etc.), tests ride primarily on teachers' assessment knowledge and abilities. Second, tests function as a platform for useful discussion for teacher education for novice and experienced teachers alike. Being aware of good practices and making every effort to ensure tests are truly useful and meaningful for our students is an unavoidable responsibility for any teacher. Third, the effects of tests on instruction and curriculum, which is widely discussed in assessment literature as 'washback effect,' will guide teachers for any changes and/or adjustments needed.

Among various tests used in language classes, this paper specifically focuses on achievement tests. Achievement tests measure what a learner has learned or achieved during a course of instruction, typically in a form of mid-term or final exams as well as chapter/unit exams. However, the study focuses on 'final' achievement tests, not 'progress' achievement tests (Hughes, 2003). The final exam was chosen for analysis because it is a prototype summative test that concludes the work of a semester, as opposed to unit/lesson or chapter tests in which only focused materials (e.g., particular

grammar patterns and vocabulary words taught in a specific unit) are tested. In addition, among three main domains of a test (i.e., purpose, content, method), this paper will specifically focus on the ‘method’ domain (i.e., how to test), not the ‘content’ domain (i.e., what to test), and address issues related to understanding how our current practices have been and provide pedagogical implications for improvement.

Literature Review

There are plenty of materials and publications on language testing and assessment in general either in book format (e.g., Allison, 1999; Brown & Abeywickrama, 2010; Fulcher & Davidson, 2007; Genesse & Upshur, 1996; Green, 2014; Henning, 1987; Hughes, 2003; McMillan, 2001) or as journal articles, reflecting its significance in language education. However, it appears that empirical studies focusing on testing at the micro level (e.g., the challenges of developing unit tests or final exams) as part of achievement tests are relatively scarce. One notable study is Barrette (2004). By analyzing 13 tests used in the beginning and intermediate college Spanish classes employing Bachman & Palmer’s (1996) “Task Characteristics” as an analytical tool, Barrette (2004, p. 58) discusses various problems detected and reported by the test authors and reviewers. She reports that the results of the analysis “reveal patterns of inappropriate input and inadequately specified procedures, tasks, scoring criteria and expected responses.” Since this study examines issues of test writing in a classroom setting from the teachers’ perspective, the findings are useful and informative to other language teachers, and thus serves as the basic model for the current study.

Considerably more studies have been conducted on KFL assessment in Korea than in America. Chang (2011) reviewed research on Korean language assessment over the past three decades, based on 146 publications from 1981 through 2011, including degree thesis (MA and Ph.D.), journal articles and books and special project reports on the topic. This review shows that research on Korean language assessment is largely categorized into 1) assessment in general, 2) proficiency tests, 3) TOPIK (Test of Proficiency in Korean), 4) achievement tests, and 5) other assessments (e.g., assessments of Korean for Academic Purposes, web-based assessments, performance assessments). One of the main findings of this study is that studies on proficiency tests far outnumber those on achievement tests. Those few studies on achievement tests focus mostly on effectiveness of a specific institution's assessment tools, analysis of current practices of various institutions, or ways to assess specific language modality such as writing. Choi (2006) is one of these studies, which compiled and compared placement tests as well as achievement tests of four major Korean language institutes affiliated with Korean universities, aiming to closely examine the current practices of these two types of tests and provide suggestions for better practices.

Those studies published in North America on testing/assessment in general and achievement tests in particular are indeed scarce. According to Choi et al. (forthcoming), which reviewed journal publications on assessment since 2000, proficiency tests, especially oral proficiency tests (e.g., Cho, 2004; Kang & Kim, 2000; H. Kim, 2016; M. Kim, 2000; Y. Kim, 2004) and placement tests (e.g., Lee et al., 2002; J. Lee, 2016; Y. Lee, 2000; Shin & Lee, 2014; Sohn & Shin, 2007) received the most attention from researchers. Notable exceptions to these include Choi (2000),

Shin (2014), and Yoon et al. (2018). Choi (2000) investigated rater reliability on the speaking portion of a unit test at the Defense Language Institute. While this is a small-scale study that was published in conference proceedings conducted at a special school like DLI, it is worth mentioning since it brings our attention to a unit test. Based on a ten-item survey questionnaire filled out by 22 faculty members along with follow-up interviews with a few of them, Choi concluded that teachers should be provided with more informative training sessions in order to ascertain rater consistency and reliability.

On the other hand, Shin (2014) examined students' perceptions of class assessment and feedback from learners' perspective in relation to the assessment process. Seventy-one learners of Korean in 2nd and 3rd year classes that he surveyed showed that they were all very aware of the importance of the assessment and feedback and that they preferred that the practical use of their language skills be assessed to the assessment of the level of grammatical knowledge or cultural knowledge.

Most recently, Yoon et al. (2018) used actual achievement test samples (i.e., vocabulary quiz, lesson test, mid-term exam, final exam) collected from six institutions in Korea and six institutions in America. They looked at the characteristics of items such as the types of questions used in these tests by skills (e.g., reading, writing, listening). The study was descriptive and example-driven rather than quantitative and trend-finding. Nonetheless, it is encouraging that studies looking closely into achievement tests are emerging.

There are a few ongoing studies that need our attention. For example, Shin and Lee (forthcoming) examined the validity and effectiveness of item content or test format in actual exam samples in KFL classes. Wang (forthcoming)

discussed teachers' beliefs and views of assessment and their current assessment practices based on in-depth interviews. Despite such recent efforts, Choi et al (forthcoming) reports that only 3% of all studies published in KFL-related journals they compiled for the past two decades were on assessment. This clearly suggests that the topic of assessment (broadly defined) is seriously under-researched. The current study is intended to begin to fill this research gap and expand the existing body of literature on assessment/testing in KFL in a classroom setting in particular.

The Study

3.1. Data & Instrument

In order to learn about current testing practices, analyzing the tests themselves was deemed to be an invaluable step to take. To that end, the heads or directors of 11 randomly selected Korean language programs in institutions across America were contacted by the researcher for samples used in the beginning (1st year) and intermediate (2nd year) Korean language classes at their respective institutions. These lower-level courses were targeted because their primary instructional goals are laying a strong foundation and equipping students with basic grammar and vocabulary on daily topics. The pedagogical approaches as well as textbook formats and structures are very similar in these levels across institutions, unlike upper-level (3rd year and beyond) courses, which vary to a great extent by instructor and program. It was assumed that comparison would be straightforward.

Two schools were unable to share their samples for 'sensitivity' reasons, two schools were unresponsive to repeated requests, and one school was disqualified (i.e., no final exams were administered). Samples of six schools were thus analyzed for each level, with 12 total tests as a final data pool.

Among those participating schools, five were located in the East and one was in the Midwest. Five schools were private schools and one was a state school. The results of the comparisons of six final exams for the 1st year and six final exams for the 2nd year are discussed in this paper. The final exam takes up 15-25% of the entire course grade in many schools, a relatively heavy weight as a single category in proportion to other requirements (e.g., attendance and participation, presentations, quizzes, homework).

3.2. Research Questions

Three research questions were posed as follows:

Q1: Are the tests similar to or different from each other in terms of the relative importance of each skill, format, length of input, and other aspects across the institutions and programs?

Q2: Are there any major differences across levels between the beginning and the intermediate levels?

Q3: What are the most common issues/problems that emerged from the analysis from a test writing perspective?

Bachman and Palmer (1996, pp. 49-50)'s "Task Characteristics" was used as an analytical tool. These characteristics include a) characteristics of the setting (physical characteristics, participants, and time of task); b) characteristics of the test rubrics (instructions, structure, time allotment, scoring method); c) characteristics of the input (length, format, language of input); d) characteristics of the expected response (format, language of expected response); and e) relationship between input and response. Among these five characteristics, characteristics of the test rubrics and characteristics of the input are discussed here because other information was unavailable through the data collected.

Findings

In reporting the findings, the characteristics of the rubrics will be discussed first and then the characteristics of the input.

4.3.1. Characteristics of the Test Rubrics

In this section, four aspects will be analyzed and they include a) language of instructions, b) structure, c) time allotment, and d) scoring method.

4.3.1.1. Language of Instructions

The instructions have to do with the language (native or target or both), channel (aural or visual or both), and specifications of procedures and tasks. As Table 1 below demonstrates, the instructional language is predominantly the learners' native language, regardless of the level. At the beginning level, all six schools (100%) provided instructions in English while four schools (67%) did the same in the intermediate level. Between these two schools which did not provide instruction in English, one school used the students' target language (i.e., Korean) and the other school used a mixture of both learners' native language, English, and their target language, Korean, for different questions. The instructions are not highly complex, as they are quite standard questions one can expect to see in a typical lower-level foreign language test. While teachers may not want the lack of understanding and/or possible misunderstanding of the directions to impede the performance of the students and thus used learners' native language exclusively in the beginning level, having instructions in the target language can indirectly test reading comprehension along with those test items the directions pertain to. Table 1 displays the result.

Table 1. Instructional Language

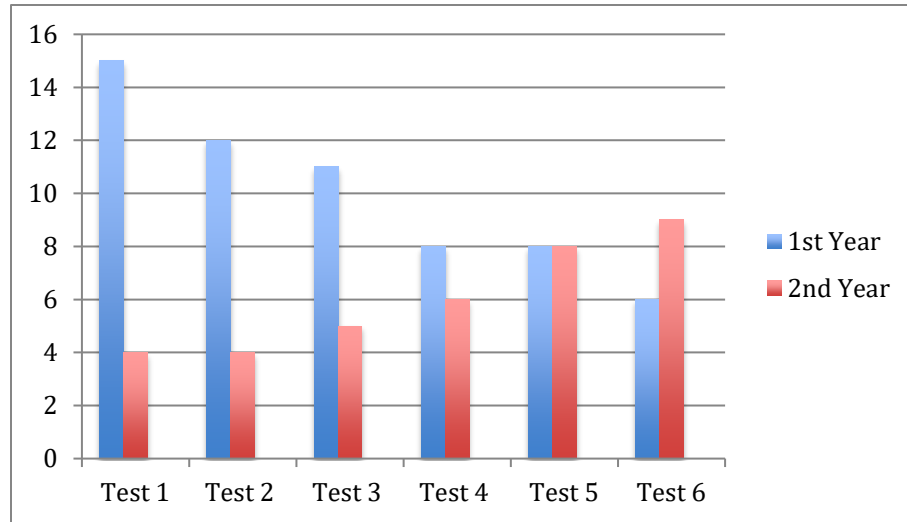
Level	NL	TL	Both	Mixed
1 st	6	0	0	0
2 nd	4	1	0	1

(Note: NL-learners' native language TL-learners' target language)

4.3.1.2. Structure

The structure includes number of parts/tasks, salience of parts/tasks, sequence of parts/tasks, relative importance of parts/tasks, and number of tasks/items per part. In regards to the number of parts or tasks in the test structure, the number ranges from 6 to 15 in the first-year test, and from 4 to 9 in the second-year test. The range is larger in the first-year than in the second-year. Two samples have 8 parts or tasks in the first year, and two samples have 4 parts or tasks in the second year, while the number of parts or tasks in all other samples varies. A general trend seems to be that there are considerably more parts in first-year than in second-year tests.

Figure 1. Number of Parts/Tasks in Test Structure



1st Yr.: 1 test (15), 1 test (12), 1 test (11), 2 tests (8), 1 test (6) – ranges from 6 to 15

2nd Yr.: 2 test (4), 1 test (5), 1 test (6), 1 test (8), 1 test (9) – ranges from 4 to 9

With respect to the salience of parts/tasks, different parts of the test are clearly distinguished from one another in almost all the tests in both the first-year samples and the second-year samples. Only one test in each level did not show any clear distinction. When the sequence of parts or tasks is analyzed, it is revealed that the test begins with listening and ends with writing in the majority of the samples. In the case of 1st year, three tests (50%) followed the order of listening-> vocabulary, expression, grammar -> reading -> writing. The other two followed a slightly different order, although keeping the pattern of beginning with listening and ending with writing. One test did not have listening or reading components. It was comprised of vocabulary, grammar, and expression (e.g. more than a single word). The second-year tests also followed a very similar pattern as the 1st year tests as the following tables display.

Table 2. Sequence of Parts/Tasks (1st Year)

	L>VEG>R> W	L>VEG>R>G> W	L>R>VEG>L> W	VEG
	50%	17%	17%	17%

(Note: total percentage is over 100 because of roundup.

L-Listening VEG-Vocabulary, Expression, Grammar R-Reading W-Writing)

Table 3. Sequence of Parts/Tasks (2nd Year)

	L>VEG>R>W	L>R>VEG>W	VEG>W
	33%	50%	17%

(Note: L-Listening VEG-Vocabulary, Expression, Grammar R-Reading W-Writing)

If both listening and writing were included in the test, virtually all tests follow a common pattern, beginning with listening and ending with writing.

The relative importance of parts or tasks in terms of score allocation for each skill presents a different picture as the following table demonstrates.

Table 4. Weight on Each Skill across Program and Level

Level	Listening	Reading	Writing	Grammar
1 st Yr.	0~21	0~30	0~22	44~100
2 nd Yr.	12~23	14~34	14~33	22~54

(Note: numbers are in percentage.)

At the first-year level, grammar takes up the largest portion of the test while other skills such as listening, reading, and writing are weighed roughly equally probably because Korean language is structurally very different from English and thus grammar was emphasized. On the other hand, at the second-year level, the relative weight of grammar decreases while the weight of writing slightly increases.

When these results were broken down by test and looked at more closely, the following results were found as displayed in Table 5 below. Test #3 is notable in that the entire test was about grammar. Two tests (#3 and #6) did not include listening in the final exam. As for reading, three tests were similar at around 20% with two outliers at 10% and 30% respectively. As for writing, one test (#1) was notably higher at over 20% while the majority (i.e. four tests) were approximately around 10% range. As for the grammar portion, three tests were around 60%, two tests around 50% with one exception at 100%. However, grammar was given the heaviest focus in all the test samples. The relative proportion of different skills was determined by the possible total scores assigned to each skill, and this was clearly marked in the test papers.

Table 5. Relative Proportion of Four Skills for 1st Year

	Test 1	Test 2	Test 3*	Test 4	Test 5	Test 6
L	16	21	0	16	10	18
R	19	21	0	10	19	26
W	22	9	0	10	13	10
G	44	49	100	64	58	46
Total	100	100	100	100	100	100

(Note: L-Listening, R: Reading, W: Writing, G: Grammar (including vocab. & expression)

(Note 2: numbers are in percentage.)

The results of the second-year tests display a very similar pattern as the first-year test where grammar was the largest proportion in all the tests except for one (Test #4). Test #3 was almost completely on grammar with a very small reading component. This test did not include listening or writing. However, the relative proportion of grammar in comparison to other skills is smaller than in the first-year tests. These findings strongly suggest that grammar is still a crucial component of instruction in the second-year courses, although slightly less so than in the first-year courses. This might have to do with the textbooks the programs use (i.e. the vast majority of the Korean programs in America use the same textbook) or the perceived importance of teaching and testing grammar on the part of the instructors.

Table 6. Relative Proportion of Four Skills for 2nd Year

Skills	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
L	17	23	0	17	12	0
R	28	34	8	28	14	11
W	17	14	0	33*	20	15
G	39	30	92*	22	54*	74*
Total	100	100	100	100	100	100

 (Listening: 0~23

Reading: 8~34

Writing: 0~33

Grammar: 22~92)

4.3.1.3. Time Allotment

None of the tests specifies the estimated time required to complete each part or task, which is common in classroom tests, unlike standardized tests.

4.3.1.4. Scoring method (how numbers are assigned to test takers' performances)

The scoring method is basically associated with the explicitness of criteria for correctness and the procedures for scoring the responses. The results showed that first, all the tests except for one test in each level clearly indicated the total points assigned to each section. Second, a few tests did not clearly indicate the total possible points for the entire test. Third, a few tests clearly stated “2 points each unless specified otherwise” or “1 point each, 10 points total” or use an expression like ‘1x4=4 points’ for each subsection of the test. Finally, explicit criteria for correctness were missing in almost all of the tests, especially in answers of limited production and extended production.

4.3.2. Characteristics of the Input

Three aspects are considered for the characteristics of input: length, format, and language of input. Table 7 below displays the results.

4.3.2.1. Length

Table 7. Length of Listening Passages (in number of words)

Level	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
1 st yr.	169	533	N/A	266	151	44*
2 nd yr.	189	342	N/A	334	839	N/A

As the table shows, the length of listening passages greatly varies ranging from 44 words to 533 words in the first year, and from 189 words to 839 words in the second year, when listening was a part of the test. Test #1 with 533 words and Test 6 with 44 words in the 1st year and Test #1 with 189 words and Test #5 with 839 words were outliers in each level. The norm for the first year seems to be between 150 and 270 words, and around 350 words for the second year. Even though it is hard to make any generalizations due to the number of samples compared, three trends clearly emerged from these data. One is that length significantly varies from test to test. Also, more words are used in second-year tests than first-year tests in general, which is natural given the level difference. Moreover, some schools do not make listening part of the final exam (#3) regardless of the level, while other schools (#6) include listening in the first-year test but not in the second-year test.

When comparisons were made at the same school between first-year and second-year tests, two tests (i.e. Test #1 and Test #4) show a slightly increased number of words used in the second year, while two tests (i.e. Test #2 and Test #5) display an erratic pattern. The number of words used in the second-year test was considerably lower than that of the first-year test in Test #2 (533 vs. 342) and the number of words used in the second-year test was considerably higher than the first-year test in Test #5 (151 vs. 839). In the case of Test #3, listening was not included in the final exam and in the case of Test #6, listening was included only in the first-year test.

Table 8. Length of Reading Passages (in number of words)

Level	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
1 st yr.	387	363	N/A	322	561	299
2 nd yr.	440	535	89	306	441	318

Similarly, the length of the reading passages varies, ranging between 299 words and 561 words in the first-year samples and between 89 words and 535 words in the second-year samples. The norm for the first year seems to be between 320 and 390 and that for the second year is between 300 and 440. What is interesting to note is that there is a considerable range among schools, but that there is no notable difference between levels. While differences between schools is understandable, no difference between levels is somewhat

perplexing. In particular, two tests (e.g. #4 and #5) show that reading passages are longer in the first-year test than in the second-year test. The sheer length of the test might not be an indicator for a ‘good’ or ‘bad’ test, but it certainly raises questions.

4.3.2.2. Format

With respect to type, there are three major test types: selected response (multiple choice tasks), limited production (form a single word/phrase to a single sentence or utterance), and extended production (longer than a single sentence or utterance). The length of the expected response was either one word or one sentence.

When item type was analyzed, examples of selected response (SR) were almost all in multiple-choice format although some responses require filling in the blanks with the most appropriate words (with or without cues), and circling/choosing the most appropriate particles/question words/answers to questions (from the box/choices/examples given), choosing the incorrect sentence, and choosing true or false. Examples of limited production (LP) include completing the conversation (with or without cues), translating the given sentences (with or without specific instructions), answering the questions in complete sentences, identifying errors and correcting them, rewriting the sentence as instructed (e.g. using intimate style ending or proper honorifics), and creating a sentence using provided expressions. Examples of extended production (EP) include writing essays and translating an entire passage. The following tables show a comparison of item types for first year and second year.

Table 9. Ratio among Response Types for VEG Section (1st year)

Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	
SR	23	27	67	52	15	100
LP	77	73	33	38	85	0
EP	0	0	0	10	0	0
Total	100	100	100	100	100	100

(Note: Numbers are percentages.)

For the first-year test, all the tests except for one test (i.e., Test #4) did not have extended production formats. All the questions in one test (i.e., Test #6) were selected response types. The majority of the tests show a mixture of selected response and limited production types, although the proportion of limited production is significantly higher than selected response type in three tests (i.e., tests #1, #2, and #5). In the remaining two tests (i.e., tests #3 and #4), there were more selected response type questions than limited production type questions.

Table 10. Ratio among Response Types for VEG Section (2nd year)

	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
SR	6	69	44	18	34	100
LP	94	31	56	82	66	0
EP	0	0	0	0	0	0
Total	100	100	100	100	100	100

As for the 2nd year test, no test had extended production type. In one test (Test #6), all the questions were selected response type. Except for one test (Test #2), there were more limited production than selected response, predominantly more in two tests (tests #1 and #4) and notably more in two other tests (Tests #3 and #5).

4.3.2.3. Language of Input

This was done almost exclusively in Korean. If there were a section on testing ‘culture notes’² of the textbook, language of the input and expected response could have been English as a way of comprehension check given how ‘culture notes’ are written. The primary purpose of this section is to provide information on a chosen topic of Korean culture (e.g., seaweed soup for birthday, traditional Korean house ‘hanok’ and ‘ondol’). However, no test samples had questions about culture. A few tests had translation questions for which the language of input and/or expected responses was English.

Summary and Discussion

Let us recall the research questions set forth at the outset of this paper. For the sake of efficiency and relatedness of both questions, research questions 1 and 2 are discussed together. They have to do with similarities and/or differences across program or level, and the following results have been found. First, the majority of test samples analyzed included listening, reading, and writing in the final exams clearly marked/labeled while a few test samples focused exclusively on grammar and vocabulary.

² Some textbooks that are commonly used in Korean programs, *Integrated Korean* series published by the University of Hawaii Press, for example, have a section called ‘culture’ in each unit. It explains relevant topics (e.g. ‘Korean housing’ when the characters talk about where they live in the unit) to the theme of the unit in English.

Given the primary course objectives of the first two years being helping learners develop basic communication skills by building a strong foundation that will enable the learners to talk about daily activities after acquiring literacy in Korean, a heavy focus on essential grammar and vocabulary is expected and thus not surprising. It appears to be a common practice in Korean programs as well as other language programs that final exams are administered in two parts—a written exam and an oral exam. Administering a speaking exam separately is unavoidable because of its format. It might also be due to the fact that oral communication has been more emphasized than written communication, especially in lower levels in general, which is in part evidenced in the popularity of the ACTFL oral proficiency interview. Unlike the oral exam, whose primary purpose is testing learners' speaking abilities, the written exam tests multiple skills typically consisting of questions testing listening (e.g. listen and answer the questions), reading (e.g. read the following passage and answer the questions), and writing (e.g. write about X). Even though the written final is divided into different sections labeled as listening or reading in almost all the tests analyzed, it is presumably for the sake of technicality and practicality. Theoretically, learners are not being tested solely for listening skills for the listening section or reading skills for the reading section in the strict sense because of the interconnectedness of these skills. They are being tested both for their comprehension or receptive skills (listening and reading alike) as well as their productive skills (i.e. writing) for all of which grammar and vocabulary knowledge is a crucial component. Then, the result of the actual assessment in the exams analyzed and reported in tables 4, 5, and 6, which is a sole reflection of distinct categories of the exams, might be different from assessing each skill separately. The question is whether each

skill can be exclusively assessed in the current practices. The analysis showed that the distinction of four skills from a practical perspective, however arbitrary it is, is still practiced in the final exams.

The discrete assessment of learners' grammar knowledge that is typically tested in a multiple choice format has been unfavorably viewed and criticized not only in foreign language education in general (e.g., Hadley, 2001) but also in KFL in particular (e.g., Kim, 2010) while 'integrative performance assessment' has been strongly advocated by researchers since communicative language teaching has gained enormous popularity several decades ago. Such shift naturally fueled a debate over whether grammar should be tested separately or embedded in the test of each skill (i.e., test efficacy), although a clear division might not be always easy or even possible. In spite of the recent popularity of integrative assessment, the need for discrete assessment cannot be denied when it comes to achievement tests, as each test aims to ensure the learning of specific grammar patterns and vocabulary words in the 'focused or assigned materials' at hand. Learners should learn what they are supposed to learn for the term and teachers are supposed to measure whether learners learned (i.e., know) the material. For instance, a very common form of assessing particle acquisition in KFL is asking students to select a correct particle in a given sentence. If students choose a correct particle, they earn that point and it is interpreted that they 'acquired' and 'know' the grammar.

However, accepting the need for a discrete assessment format and performing integrative assessment need not to be mutually exclusive as they serve different purposes. Discrete assessment is most effective for testing acquisition of a specific

grammar point as well as for placing students for a specific skill development course. If the need for discrete assessment were granted, it would be ideal to properly combine discrete assessment and integrative assessment for different purposes at different points in time in the entire learning process. One suggestion would be creating unit tests as discrete assessments and final exams as integrative assessments (Wang, 2010), in which case the format and focus of unit tests and final exams should be properly differentiated.

Second, the relative proportion of the four skills notably varies from school to school, while it varies less between levels. Again, the grammar section displays a conspicuous trend. Irrespective of school and level, its proportion is considerably higher than other skills, and this pattern is more salient in the first-year tests than in the second-year tests, reflecting not only how important grammar and vocabulary instruction is viewed in the beginning stage of learning but also how much of classroom instruction time is spent on grammar teaching, assuming that heavy focus reflects the amount of instruction. The range for other skills (i.e., listening, reading, writing) was relatively small. According to Choi (2006), who compared several institutions in Korea, four skills are weighted equally or almost equally in all the institutions that she surveyed, although two institutions assign different weight for different skills based on the level (lower levels vs. higher levels). For instance, the proportion of the four skills in one program is 30 /30 /20 /20 (listening/ speaking/ reading/ writing) for levels 1-3, but equally weighted (25/ 25/ 25/ 25) for levels 4-6. The structure of final exams used and reported in Choi (2006) and those used in this study is not exactly comparable so should be taken with caution, but Choi's findings shed some light on the proportion of four

skills. This leads to a question as to what a proportionally ideal exam should be like for each level. Should the proportion of each skill differ as the level goes up, and if so, on what grounds? And where does grammar fit in?

As we recall, the length of listening ranged between 44-533 in the beginning level and 189-839 in the intermediate level and that of reading ranged between 299-561 in the beginning level and 89-535 in the intermediate level. It was the case that the number of questions for the given reading or listening prompts did not significantly vary regardless of the length of the passages. One may wonder how many questions a test writer can create from these listening or reading passages. This may depend on the difference between the two skills, the task complexity (e.g. finding main idea vs. finding details) as well as type of learners (e.g. heritage learners vs. non-heritage learners). For example, the processing cost of listening comprehension is higher than the processing cost of reading comprehension for non-heritage learners, whereas it would be the opposite for heritage learners. A high level of listening comprehension and low level of reading comprehension is the profile of a typical heritage learner who grew up in a Korean household. What this implies is that various factors that would affect assessment need to be considered for the 'ideal' distribution of each skill. This could pose a challenge for teachers to write a test for a mixed-class of heritage and non-heritage learners because many Korean programs do not run separate tracks except for a few large programs.

It was reported in the findings section that the relative importance of four skills in terms of task proportion does not vary considerably across levels if the two outliers that display a somewhat extreme tendency (i.e., no listening or writing,

exclusively on grammar) are disregarded. Choi, citing Kim et. al. (1993), suggested that speaking and listening should be given more weight at the beginning level, and more weight on reading and writing in the intermediate and advanced levels. In other words, it was recommended that the assessment emphasis should be placed on spoken language for lower levels and on written language for higher levels, implying that level was a factor to be taken into account. While there is no right or 'set-in-stone' guidelines to follow, the findings suggested each program to revisit this issue from both theoretical (i.e., what would be most beneficial for their learners) and practical (i.e., the convenience of test administration) perspectives.

Third, the sequence of parts/tasks is found to be very similar in all tests regardless of program or level, which turned out to be one of the rare common characteristics of the analysis. Except for the test that only tested grammar/vocabulary in first year and the test that did not include listening and reading in second year, all the remaining test samples placed listening in the beginning and writing at the end of the test. This seems to be a 'standard' practice that is perhaps motivated by convenience or practicality of test administration. Listening prompts should be played to the entire class before students move on to other sections, for which they could be on their own pace, and thus it is reasonable to administer this section at the beginning of the test. On the other hand, writing is usually an open-ended free production task and is subject to various factors (e.g., test takers' content formation, language proficiency, speed and amount of production, among others) so that more room and flexibility is allowed for. Thus, it is placed at the end of the test. In terms of the test structure, however, the range of number of parts/tasks is larger in the first-year tests than in the second-

year tests, although there is no notable difference in terms of salience of parts/tasks, and sequence of parts/tasks. First-year tests are broken down in more pieces/sections than those of second year probably because there are more ‘small’ grammar points to be checked such as particles and verb conjugations that are typically tested in a certain format (e.g., “choose the correct particle”) to be effective.

Fourth, selected response (i.e., multiple choice) and limited production (i.e., short answer; from a single word or phrase to a single sentence or utterance) are the most popular response types in the grammar/vocabulary section of all the samples analyzed. There are virtually no extended response (i.e., longer than a single sentence or utterance) questions. The ratio is notably higher for limited production than for selected response in both first- and second-year tests in the majority of test samples, showing no difference by level, with one outlier in which only selected response was used for both first- and second-year tests. Typical limited production questions are, ‘fill in the blanks with the appropriate words,’ ‘complete the following sentences/conversations with cues provided,’ ‘answer the following questions with the grammar patterns provided,’ ‘identify errors and correct them,’ or ‘rewrite the sentence as instructed.’ Each item type has its own advantages and disadvantages³ and thus it would be prudent that teachers carefully determine the most appropriate test type while being mindful of the purpose of assessment of each task. Here again, an integrative format could be adopted on the grounds that it is final exam, a prototype summative test. For example, by

³ For a list of advantages and disadvantages of each test type, please see <https://www.facultyfocus.com/articles/educational-assessment/advantages-and-disadvantages-of-different-types-of-test-questions>

making limited production questions in a sequence (i.e., discourse-based) rather than one or two unrelated sentences in isolation, teachers will be able to assess learners' knowledge as well as skills from a communicative perspective.

Fifth, the length of inputs measured by the number of words in the reading and listening passages notably varies from program to program. For level differences, the results are hard to generalize, since they do not display any dominant pattern. However, it is apparent that the length of input in both listening and reading passages does not necessarily increase in second-year tests as one would have expected. It slightly increases in one test and drastically increases in another for the listening input. Similarly, it slightly increases in two tests, and notably increases in one test for the length of the reading input. One might think that learners at the second-year level should be able to handle longer and more complex texts as well as tasks, and the tests should properly measure learners' ability to handle such complexity. The findings, however, do not seem to support such reasoning and the question of level articulation is called for. Level articulation is a critical issue in language program management both from an instructional perspective as well as an assessment perspective, which are in fact closely interrelated. The majority of Korean language programs in U.S. colleges and universities use the same textbook for the beginning and intermediate levels as mentioned before⁴ and are presumed to practice a similar instructional approach, yet the findings of this study exemplified in the length of input for instance indicate the lack of articulation at the micro level.

⁴ According to AATK survey conducted in 2007-2008 academic year, approximately 70% of the schools surveyed uses *Integrated Korean* for first year and second year.

More communication between level instructors is recommended for making informed decisions.

As for research question 3, the most common issues/problems that emerged from the analysis from the perspective of actual test writing, several common pitfalls in test writing surfaced. Barrette (2004, p. 63) lists the most common problems in writing/developing tests: the specifications of procedures and tasks, the relative importance of each section and item, the explicitness of criteria for correctness, and the length of the input. The specifications of procedures and tasks and the explicitness of criteria for correctness are further discussed below.

(1) Specifications of procedures and tasks

A common problem within this framework is the ambiguous specifications of procedures and tasks. More specifically, test writers “omitted information that might seem obvious to the test authors although not to examinees” (Barrette, p. 63). In one test in the sample, the instructions read, “Identify the errors in the following sentences and correct them. Each sentence contains only ONE error.” The sentence is *Minji-nun mayil kelulo kongwen-ey ka-yo* (민지는 매일 걸으러 공원에 가요), meaning ‘Minji goes to a park for a walk every day.’ However, there was no information regarding whether there will be any penalty for misidentification and/or miscorrection. The intention of the test writer must have been testing whether his/her students distinguish between AVst (u)le ~ey kata/ota (AVst (으)러 ~에 가다/오다) meaning ‘to go to in order to do Verb’ as opposed to N (u)lo kata/ota (N (으)로 가다/오다) meaning ‘to go by means of N’. Teachers know that this is a commonly made error by students. The

instructions did not specify whether these errors were grammar-related or vocabulary-related. Will this information be necessary or is it better to leave intentionally ambiguous so that students should figure out? What exactly is being tested here?

In questions related to listening, there was no specification as to how many times the passage(s) and/ or question(s) will be repeated, unlike most standardized tests; no specification on whether the test takers will listen to the tape or to the test giver in person, if it matters; no specification on whether the answers should be written in full sentence or not (e.g., instructions read “Listen to the question and answer them in Korean.”). Some students may write their answers in full sentences that would require more time than those who simply write the key words. Another problem with the writing section of the test is that the degree of specification seems to be inconsistent. Some tests are not sufficiently specific while other tests are overly specific. Examples vary from “An open-ended essay on an assigned topic with specified expected length” to mandatory elements required is specified (e.g., specific grammar patterns, particles, etc.) as opposed to saying, “Utilize as many grammar patterns we have learned as possible” to “Use the past tense/future tense/honorific expressions at least 3 times. Each sentence should have at least 4 words. Underline all the words that meet these requirements.”

(2) Explicitness of criteria for correctness

As Barrette pointed out that “Omission of minimal requirements for full credit, availability of partial credit, or length, type, or language of the expected response,” omitting the explicitness of criteria for correctness was very common

(Barrette, p. 64). This was especially common in open-ended limited production and extended production responses (e.g., complete the following dialogue, translate the following sentence or paragraph), which appeared to be most problematic in the writing section of a test. Some examples in the writing section in which criteria for scoring are absent, vague or insufficient are as follows:

- (1) Write at least 12 sentences about your winter break following the instructions provided below (xx pts.)
- (2) Give a narration about yourself. You should include at least the following information..... (xx pts.)
- (3) Write about both of the following topics in two paragraphs. Use at least (xx pts.)
- (4) Write a short essay about one of the following topics in a minimum of 12 sentences. Use the deferential ending only. The requirements of this essay are (xx pts.)

There were also other rare and minor issues such as inaccurate instructions (e.g., from the box – but no box) or typos that can be easily corrected by proofreading the exam.

Grammar and listening sections are relatively straightforward (discrete?) and pose fewer problems in grading. Speaking and writing sections, however, are less transparent and thus more challenging with respect to scoring, due to their openness as well as lack of objective criteria. Speaking is assessed separately by a final oral test in many programs, for which detailed assessment rubrics must be used. On the other hand, writing is part of the final exam and no explicit scoring criteria or guidelines are provided in the test

itself other than the total points one can get, which of course does not mean that criteria do not exist or are not used. Teachers might have communicated with the students before the test, but that is unknown. What is suggested is to clearly communicate with the learners what the main purpose of such a writing test is; whether it tests usage of certain grammar patterns and/or certain vocabulary words and expressions that are taught in the semester (i.e., linguistic knowledge and production); whether spelling errors will be counted and penalized; and whether idea flow or organization is assessed (i.e., metalinguistic aspects) and so on. Although it appears to be the case in sample tests used in the current study that writing is part of the final exam, testing writing separately might be an idea to pursue to increase the effectiveness, given the nature of writing.

Here Davidson & Lynch (2002)'s test specification format would be useful for the teachers to bear in mind when writing tests and to apply where relevant. Its four components are general description, prompt attributes, response attributes, and sample items. The general description describes the core of the learning objective. Prompt attributes describe what the student will encounter. Response attributes describe the way in which the student will provide the answer. And the sample item is the sort of item or task this specification should generate.

There are many challenges that teachers face and barriers that they struggle with in conducting general assessment duties as well as developing specific tests as discussed in Yoon et al. (2018) and Wang (forthcoming). They are wide-ranging, some of which include mismatches for gaps between students' daily performance and assessment (i.e.,

tests) and how to deal with it, assessments of intercultural competence and/or pragmatic competence, gaps between students' learning goals set at the beginning of the semester and learning outcomes at the end of the course, establishing assessment criteria for open-ended questions, dealing with students' diversifying L1, heavy emphasis on vocabulary and grammar, and so on. Some of these challenges are general in nature while others are more directly associated with summative tests (i.e., final exams). Since assessment is a very big topic that encompasses so many related aspects, this is a small step taken by looking closely at the current practices of final exams.

Limitations of the Study and Implications for Teaching

While this study may be one of the very few studies in KFL, if not the only one, that investigates current testing practices of Korean language courses in colleges and universities in America, it is not without limitations. The study is exploratory with a primary objective of learning about how testing is done and what the common issues are across institutions (i.e., fact finding and attention gathering). Its analysis is based on written final examinations with the understanding that for many programs the final examination serves as the culmination of all the learning taken place during a given semester/term. Analysis of other tests (e.g., unit test, mid-term exam) and comparing them with final exams might present different results, although it is suspected that the main issues may remain the same, since final exams are generally understood as extended versions of unit tests. Also, despite the fact that five or six tests for each level still demonstrate some interesting trends, more samples would have been desirable as they could have yielded results with stronger generalizability. Due to difficulties in collecting

the data, participation rate was lower than was hoped. Institutions also seem to exercise different policies on being open about the tests. The researcher's personal communications with a few colleagues have revealed that some teachers find sharing tests with others uncomfortable for various reasons. They consider test writing a 'private' practice, a perception that needs to be deconstructed. Given the size of the Korean language program in many schools where it is run by a faculty member single-handedly or by a couple of faculty members who are in full charge of their respective courses, receiving feedback on the developed tests and/or collaborating in the process of test development has not been a common practice. As advocated in Brown and Hudson (1998), the idea of using multiple sources of information is helpful in making assessment-related decisions. Good test writing practices indeed require a collaborative effort and cycles of feedback.

Since the focus of investigation of this study is on the various aspects of test writing and administration as clearly stated in the Introduction, not on a close examination of items of the final exams, a content analysis needs to be done in future research.⁵ Although what will be tested in terms of content is predicted in achievement tests such as final exams, unlike other types of tests (e.g., proficiency tests, placement tests, etc.), an analysis of content validity would provide some valuable insight for the teachers. In addition, a 'washback' effect needs to be evaluated by looking into the relationship between instruction and assessment (i.e., comparing the course objectives against learning outcomes and how learning outcomes are measured through the format of the test), which

⁵ Shin (2016) and Shin and Lee (forthcoming) partially did item analysis based on actual classroom test samples.

is an ultimate goal of the assessment. Furthermore, a counter check between the syllabus and the assessment would be ideal. Some teachers might state how tests are administered or how different skills are tested in their syllabus. Some teachers might provide detailed information as to how tests will be done separately. Looking into any communication between teachers and students will also shed some light on the issue.

The findings of this study lead us to make some practical suggestions especially from logistical aspects of test writing and administration (focus on form). It is desirable that teachers determine the relative weight for different skill modalities (i.e., listening, reading, writing and speaking) for different levels ideally, informed by theoretical foundations and such reasoning should guide the proper proportion of the tests. Likewise, task difficulty informed by inherent complexity of items tested through error analysis, for example, should be reflected in the form of test/item type (i.e., selected response vs. limited production response vs. extended production response) for different levels as well as for different questions. Finally, teachers are encouraged to think from the perspective of test takers to help them design clearer and less ambiguous tests.

Within one's own institution, seeking feedback from and ensuring follow up discussion with colleagues (if one has a colleague) when a draft is ready would be enormously useful. Despite the fact that teachers have their own beliefs, preferences, and philosophy in regard to testing and assessment, open discussion and close communication with someone equally qualified during the entire process of test planning, developing, and writing would be highly desirable. At the level of the field in general, on the other hand, focused

workshops for teachers should be offered as part of the professional development program. Previous research (e.g., Malone, 2008; Montee et. al., 2013) reports that many foreign language teachers, especially those in less commonly taught languages, do not have proper training or have limited training in assessment including test development or writing. They feel unprepared and have low ‘assessment literacy’ (e.g. Djoub, 2017). Teaching experience does not necessarily make a teacher a good assessor but training does, as shown in Zhang & Burry-Stock’s findings (2003, 323) that “regardless of teaching experience, teachers with measurement training report a higher level of self-perceived assessment skills” in all sorts of assessment. For many teachers, pre-service and in-service alike, they teach as they have learned. While teachers learn many things as they do at the micro level instruction, testing and assessment is something that requires formal training and constant reflection given its enormous impact on teaching and learning. More training opportunities would certainly be beneficial.

The final examination is an indispensable component of a course in any college curriculum, let alone Korean courses, serving the major purposes of assessing students’ learning for a given period of time and assigning a grade for the course as an official proof of achievement. Whether a final exam is necessary in a Korean language course is one issue (i.e., some programs do not give one for various reasons) and how to make it truly useful and effective to fulfill its intended purposes is another. If the final exam continues to hold its place that it has been holding, a serious reflection of our current practices will bring fruitful discussion. Testing – in terms of both design and grading - has been taken for granted as an important duty of teachers, and yet lack of scholarly attention to this aspect of

instruction has resulted in a research void. Needless to say, the field of KFL needs to engage more in the discussion of how to develop and administer tests that are valid, reliable, and authentic in all modalities, that better assess students' achievement in alignment with the curricular goals of the course.

Concluding Remarks

Twelve final exams of first year and second year Korean courses from six colleges and universities in the U.S. were analyzed in this study for the purpose of increasing our understanding of current practices of the field with respect to various aspects of the test as an important assessment tool for the learning of our students. Test writing and administration is ultimately the responsibility of course instructors on the low level and program heads on the high level. While developing, administering, and grading tests is something of routine in the teaching process, assessing the current situation and thereby raising everyone's awareness as to how this process can be more effective and less problematic is what this paper has intended to achieve. In addition, it is hoped that this study becomes a small steppingstone for facilitating more scholarly discourse and for encouraging further research on this topic thereby improving the understanding of this crucial aspect of teaching practice. Just as teaching is an art that requires so many combined skills, fair and effective assessment through test management is an integral part of teaching that will significantly enhance the learning experience of our students.

References

- Allison, D. (1999). *Language testing and evaluation: An introductory course*. Singapore: Singapore University Press.
- American Association of Teachers of Korean. Retrieved on January 2020.
<http://www.aatk.org/www/html/textbooks.html>
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: designing and developing useful language tests*. Oxford: Oxford University Press.
- Barrette, C. (2004). An analysis of foreign language achievement test drafts. *Foreign Language Annals*, 37(1), 58-70.
- Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices*. White Plains, NY: Pearson Education.
- Brown, J. & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32(4), 653-675.
- Chang, E. (2011). Korean language education assessment studies (Hankuke Gyoyuk Pyungka Yeongu). *Bilingual Research*, 47, 351-382.
- Cho, S. (2004). Oral Proficiency Interview: Pros and cons. *Korean Language in America*, 9, 144-151.
- Choi, E-G. (2006). The practice and task of the Korean language testing. *Journal of Korean Language Education*, 17 (2), 289-319.

- Choi, H., et al. (forthcoming). Annotated bibliography of KFL. In Y. Yu Cho (Ed.) *Teaching Korean as a foreign language: Theories and practices*. NY: Routledge.
- Choi, I. (2000). Revisited reliability on speaking portion of unit test at DLI: Rater reliability. *Korean Language in America*, 5, 225-229.
- Davidson, F., & Lynch, B. (2002). *Testcraft: A teacher's guide to writing and using language test specifications*. New Haven: Yale University Press.
- Djoub, Z. (2017). Assessment literacy: Beyond teacher practice. In R. Al-Mahrooqi et al. (Eds.), *Revisiting EFL Assessment: Critical Perspective* (pp. 9-27). NY: Springer.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. NY: Routledge.
- Genesse, F., & Upshur, J. (1996). *Classroom-based evaluation in second language education*. Cambridge: Cambridge University Press.
- Green, A. (2014). *Exploring language assessment and testing: Language in action*. London: Routledge.
- Hadley, A. O. (2001). *Teaching language in context* (3rd Ed.). Boston: Heinle & Heinle.
- Henning, G. (1987). *A guide to language testing: Development, evaluation, research*. Boston: Heinl & Heinl.
- Hughes, A. (2003). *Testing for language teachers* (2nd Ed.). Cambridge: Cambridge University Press.

- Kang, S. & Kim, M. J. (2000). Assessing levels of proficiency in Korean. *Korean Language in America*, 4, 211-222.
- Kim, C. (2010). Assessing grammatical competence of Korean language (Hankuko Munbup Nungreokoi Pyungga), *Proceedings of the 20th International Conference on Korean Language Education*, 79-88.
- Kim, H. (2016). Comparing native and non-native rater assessment of Korean oral proficiency: A FACETS analysis. *Korean Language Education Research*, 51(5), 83-113.
- Kim, M. J. (2000). Treatment of loan words in Korean proficiency tests. *Korean Language in America*, 5, 249-255.
- Kim, Y. (2004). What has language proficiency got to do with accommodation in interviewer-interviewee interaction: the case of Korean Oral Proficiency Interview discourse. *Bilingual Research*, 25, 41-62.
- Lee, D. J., Lee, Y. G., & Park, S. (2002). Web-based Korean language placement tests. *Korean Language in America*, 7, 261-273.
- Lee, J. (2016). Research on Korean language placement test development: based on selected English language placement tools. *Journal of Korean Language Education*, 27, 99-131.
- Lee, Y-G. (2000). Designing a Korean language placement test. *Korean Language in America*, 4, 181-210.

- Lee, Y-G., Yuen, S., Kim, H., Ahn, C. S., Yoon, S., Baek, S. G., and Park, S. (2002). Guidelines for item-writing for the curriculum-based placement test: development and application. *Korean Language in America*, 7, 247-260.
- Malone, M. (2008). Training in language assessment. In E. Shohamy & N. Hornberger (Eds.), *Encyclopedia of language and education* (2nd Ed., Vol. 7: Language Testing and Assessment) (pp. 429-462). New York: Spring.
- McMillan, J. H. (2001). *Classroom assessment: Principles and practice for effective instruction*. Boston: Allyn and Bacon.
- Montee, M. et al. (2013). LCTL teachers' assessment knowledge and practices: An exploratory study. *Journal of Less Commonly Taught Languages*, 13, 1-31.
- Park, S. (2001). Exploring the possibilities of WBLT for operational testing purposes: Web-based Korean as a foreign language test. *The Korean Language in America*, 6, 101-110.
- Shin, S-C. (2014). Students' perceptions of assessment, feedback, and categories. *Journal of Korean Language Education*, 25(4), 51-67.
- Shin, S-Y. (2016). Investigating the validity of KFL text difficulty as defined by the ILR Reading Scales. *Journal of Less Commonly Taught Languages*, 18, 159-174.

Shin, S-Y., & Lee, H. S. (2014). Validating a semi-adaptive web-based Korean placement test. Paper presented at American Association of Teachers of Korean annual Conference, Boston, MA. June.

Shin, S-Y., & Lee, H. S. (forthcoming). Korean as a foreign language assessment. In Y. Yu Cho (Ed.), *Teaching Korean as a foreign language: Theories and practices*. NY: Routledge.

Sohn, S-O., & Shin, S-K. (2007). True beginners, false beginners, and fake beginners: Placement strategies for Korean heritage speakers. *Foreign Language Annals*, 40, 407-418.

Wang, H-S. (2010). Discussion of Chungsook Kim's Assessing grammatical competence of Korean language. *Proceedings of the 20th International Conference on Korean Language Education*, 89-90.

Wang, H-S. (forthcoming). Classroom-based assessment practices of college Korean language teachers: A qualitative study. In M. Jeon, M. Figueredo, & F. Carra-Salsberg (Eds.), *Globally informed design and praxis in languages, literatures and linguistics curricula*.

Yoon, S-S., Kang, M-S. & Sim, J-K. (2018). An analysis of assessment method in Korean classes in Korea and America and finding ways to improve. Presented at the annual conference of the American Association of Teachers of Korean. University of Toronto, June 2018.

Zhang, Z. & J. Burry-Stock (2003). Classroom assessment practices and teachers' Self-perceived assessment skills. *Applied measurement in education*, 16(4), 323-342.