English Language & Literature Teaching, Vol. *15*, No. 2 Summer 2009

The Need for Level-based Criteria in the Assessment of Oral Proficiency

Haeyoung Kim
(The Catholic University of Korea)

Kim, Haeyoung. (2009). The need for level-based criteria in the assessment of oral proficiency. *English Language & Literature Teaching, 15*(2), 169-184.

The purpose of this study is to identify the most important factor(s) that contribute(s) to the overall oral proficiency and to examine whether the factor(s) play(s) a different role depending on the proficiency level of learners. Learners were divided into novice group and advanced group and were asked to produce an oral recording of a story based on a comic strip. The recordings were transcribed and graded by three raters. According to the results, the raters attained high inter-rater reliability when assessing advanced learners. However, the reliability level became considerably lower in the assessment of novice learners. The best predictor of oral proficiency among novice learners was sociolinguistic competence and fluency, while grammatical competence and fluency were the strongest predictors for advanced learners. The results suggest the need for a separate assessment tool for different proficiency levels and the need for a different focus in the classroom depending on the learners' proficiency level.

[oral proficiency/criteria/level/novice/advanced]

I. INTRODUCTION

While the field of language education sets teaching to achieve language proficiency as its ultimate goal, there is little agreement as to what exactly is language proficiency and how it should be assessed. Depending on the types of tests, the test-takers' ability is analyzed and often compartmentalized into different components that can be quantitatively scored by raters. Along with traditional testing on reading and writing, recent technological development has accelerated the testing of speaking which, in the past, was regarded as difficult to administer on a large scale.

In light of the growing importance of oral tests in attaining admission to prestigious schools, studying abroad, or getting a job, the popularity of these tests is growing rapidly.

As demand for oral proficiency testing increases, assessment tools have been developed to rate the test-takers' speaking ability on a massive scale thanks to the wide dissemination of test equipment using the Internet and computer technology. Tests such as TOEFL (Test of English as a Foreign Language), IELTS (International English Language Testing System), and TEPS (Test of English Proficiency developed by Seoul National University) all include oral testing components with its own set of assessment criteria. TOEFL, for example, assesses linguistic resources (grammatical accuracy, grammatical complexity, and vocabulary), phonology (pronunciation, intonation, rhythm), and fluency while IELTS breaks down oral performance into the following categories: fluency and coherence, lexical resource, grammatical range and accuracy, and pronunciation. TEPS has six criteria: pronunciation, fluency, grammatical competence, discourse competence, sociolinguistic competence, and overall comprehensibility. Each test uses slightly different terms in its criteria but basically shares commonalities, such as grammatical competence, pronunciation, and fluency.

While all the components of oral proficiency are important in the mastery of one's speech ability, it would be of educational value to explore which factor has the most weight in determining oral proficiency and to examine whether the weight differs depending on the learner's proficiency level. If, for example, novice learners face different types of problems from their advanced counterparts when speaking, identifying the differences and developing curriculum to meet their specific needs can lead to efficient teaching and learning. Thus, this study will focus on whether the factors affecting the outcome of the oral proficiency assessment differ among level of proficiency.

II. LITERATURE REVIEW

1. Holistic vs. Analytic Rating Scale

In the assessment of oral proficiency, the assessment scale (i.e., holistic or analytic) and rater reliability are the key components to consider. Currently, the assessment of foreign language oral proficiency adopts two types of scales: holistic and analytic rating scales. These two are widely used in the education environment, not only in the assessment of language proficiency but also in math, science, and even the arts.

Holistic scales offer several dimensions together while analytical scales offer a separate scale for various dimensions. Holistic assessment views the final product or performance as a whole and refuses to separate the task. Scores are based on descriptions that describe the final product in its entirety. As a result, it is a quicker and a more efficient means of assessment. While holistic assessment is widely used in writing, it is an effective tool in

other areas of language proficiency, including speaking. The downside of holistic assessment is that it fails to offer specific information on the performance of the students. As Douglas (1994) points out, a single summary score may actually represent a more complex picture of a person's proficiency level and makes it difficult to pinpoint the strength and weakness of a learner.

Analytic assessment, on the other hand, dissects the final product into categories which are each given scores. Each category is treated as an independent entity and raters must pay special attention not to have one category be affected by the other. Analytical scoring can be useful in that it offers detailed insight and more diagnostic value to the test-taker who can benefit from the feedback that is broken down into categories and thus help them focus more on their weaknesses (Iwashita & Grove, 2003).

Whether it be holistic or analytical, the adoption of a clear criteria is the key to an accurate assessment of a student's performance. Both types are used in the assessment of foreign language oral proficiency.

1) Holistic Rating Scale

The most widely known holistic rating scale is the ACTFL (American Council on the Teaching of Foreign Languages) Proficiency Guidelines (1999) which were originally published to identify the foreign language proficiency of U.S. Government Employees. The Guidelines define four main levels of language proficiency: Superior, Advanced, Intermediate, and Novice. Each level is subdivided into three levels, with the exception of the Superior level (Superior), Advanced(advanced high, advanced mid, advanced low, Intermediate(intermediate high, intermediate mid, intermediate low), Novice(novice high, novice mid, novice, low)). Today, additional language-specific guidelines are under development or exist as circulating drafts for Russian, Japanese, Chinese, Hindi, Arabic, Hebrew, Korean, Hausa, Indonesian, and a number of other languages. These guidelines have exerted considerable influence on the organization of curriculum as well as on the pedagogical approaches employed by instructors in the classroom (Thompson, Thompson, & Hiple, 1988).

2) Analytical Rating Scale

Tests that use analytical rating scales are TOEFL, IELTS, and TEPS based on categories proposed by Canale and Swain (1980) and Bachman (1990) (see Table 1).

TABLE 1

Assessment Criteria for Oral Proficiency

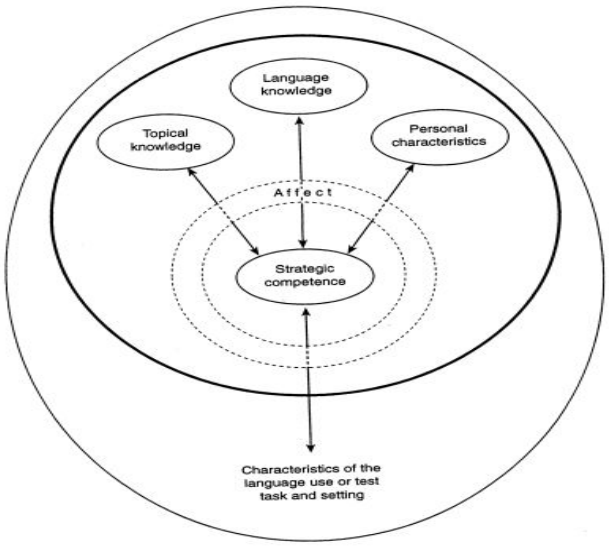
TOEFL	IELTS	TEPS
1. Grammatical Accuracy and Complexity	1. Grammatical Range and Accuracy	1. Grammatical Competence
2. Pronunciation	2. Pronunciation	2. Pronunciation
3. Fluency	3. Fluency and Coherence	3. Fluency
4. Vocabulary	4. Lexical Resource	4. Discourse Competence
5. Comprehension		5. Sociolinguistic Competence
		6. Overall Comprehension

Canale and Swain (1980) divided communicative competencies into four categories; grammatical competence, sociolinguistic competence, discourse competence, and strategic competence while Bachman (1990) offers an additional element to the assessment criteria (See Figure 1). Bachman's model includes language knowledge, topical knowledge (knowledge of the real world), and a strategic competence that can take advantage of the two aforementioned competencies as well as a psychophysiological mechanism (personal characteristics) to exert such competencies. It is also based on a hierarchical structure and shows that strategic competence executes and controls language competence and topical knowledge, rather than being a part of language knowledge.

FIGURE 1

Some Components of Language use and Language Test Performance

(Bachman **&** Palmer**,** 1996, p. 63**)**



There have been a number of studies that have examined the weight of individual criterion the overall assessment. And the outcome has yielded contradicting results. McNamara (1990), for example, found grammar and expression to be the most important factor in the assessment of overall communicative effectiveness. Adams (1980) also found vocabulary and grammar to play a key role in distinguishing proficiency levels. Brown, Iwashita, McNamara, and O'Hagan (2003), on the other hand, found that intelligibility and fluency, rather than structural accuracy, played the most important role. Higgs and Clifford (1982) went further and posited that the weight of individual factors would differ depending on the speakers' proficiency level. According to the Relative Contribution Model (RCM), vocabulary and grammar were major factors across levels, while the importance of pronunciation, fluency, and sociolinguistic factors would increase as the level went up. In a study with French learners by De Jong and van Ginkel (1992), pronunciation was the greatest contributor to overall performance at lower levels, while fluency was the determining factor at higher levels. Kim (2009) found that grammar had a significant weight in the oral proficiency score for novice learners but not for advanced learners. In a study of rating patterns in the Occupational English Test (OET), Iwashita and Grove (2003) studied whether a particular analytical criteria was more favored by raters than others. Results showed that one analytical criterion, 'Comprehension' came out as misfitting and thus did not accurately estimate the speaker's speaking ability. Despite differences in weight, however, each criterion (intelligibility, fluency, comprehension, appropriateness, grammar/expression) was eventually found to significantly contribute to the overall assessment score.

2. Inter-rater Reliability

Oral tests are often rated by two or more raters. This is to assure reliability in the assessment. If there is a wide discrepancy among the scores, then raters are obliged to discuss why they gave certain ratings. In some cases, another rater is called in to settle any discrepancy. This practice is mostly used with high stake testing such as state mandated testing.

In a study that looked at the role of raters, Wylie (1993) found high rater reliability (.906) in the assessment of 18 IELTS interviews. Kenyon and Malabonga (2001) confirmed that computer-based oral proficiency interviews could also ensure high rater reliability comparable to live interviews. On the other hand, Conlan, Bardsley, and Martinson (1994) failed to establish inter-rater reliability in audio recording compared to live interviews. This study showed that extralinguistic factors play a different role among raters. Clark (1988) also found that problems arose when multiple raters assessed oral proficiency of both live and semi-direct interviews.

Kenyon and Tschirner (2000) found differences in inter-rater reliability depending on the proficiency level of test-takers. In a study that examined interviews of 20 speakers ranging from Novice High and Intermediate Low to Intermediate Mid on the ACTFL oral proficiency interviews, inter-rater reliability was compromised in the Novice High, but not in other levels. Kim (2009) found that inter-rater reliability in the assessment of comprehensibility and grammatical competence was high for advanced learners but low for novice learners. These results imply that a different set of criteria and/or scoring techniques might be needed for raters so that they can employ different scoring strategies depending on the proficiency level of test takers.

III. METHOD

1. Participants

The study included 59 one-minute narratives, each rated and completely transcribed. A total of 59 university students (Male: 22, Female: 37) participated in the study.

2. Procedure

Based on the results of a brief one-on-one pre-interview in English, the students were divided into two groups, beginner group (Male: 9, Female: 21) and advanced group (Male: 13, Female: 16). The pre-interview was held once and lasted about 3~5 minutes each, talking about topics related to their daily lives, such as "Which courses do you enjoy the most?" "What are your plans after graduation?" "What kind of extracurricular activities are you involved in?" The study used an assessment scale of six criteria, consisting of five single-feature analytic criteria (Pronunciation, Fluency, Grammatical Competence, Discourse Competence, Sociolinguistic Competence) and one global assessment criterion, Overall Comprehension, identical to the Assessment Grid for Speaking Proficiency used in TEPS (See Appendix A).

Each group was given a comic strip consisting of eight scenes (pictures, no words included) that depicted an incident (See Table 2 and Table 3). Students were given one minute to prepare and one minute to speak. The speeches were recorded into mp3 files for analysis.

TABLE 2

Description of the Comic Strip (Novice Group)

Scene	Contents
1	A young man and a woman prepare for a trip and pack bags.
2	The man and the woman leave for the airport by bus.
3	The man and the woman wait in line at the check-in counter.
4	The woman discovers she forgot to bring her passport.
5	The woman calls home and an old lady picks up the phone.
6	The old lady finds the passport in a drawer.
7	The old lady takes a taxi to the airport with passport in hand.
8	The man and the woman bid farewell to the old lady with passport in hand.

TABLE 3

Description of the Comic Strip (Advanced Group)

Scene	Contents
1	A woman and a girl are driving in a car in the country.
2	The car suddenly shakes violently.
3	The woman finds the car has a flat tire.
4	The woman tries to call for help on her cell phone.
5	The cell phone's battery is dead. The girl looks worried.
6	The woman takes out a spare tire from the trunk.
7	The woman changes the tire. The girl is sitting beside her.
8	The woman and the girl are driving again. The girl gives the woman a thumb's up.

3. Raters

Three independent raters assessed the oral performance of both groups. Ratings were done independently by three testers who had at least five years of teaching in an EFL environment. The raters were provided with an assessment grid used in TOP (Test of Oral Proficiency in English) administered by the Seoul National University (See Appendix A). Background information of the raters is shown in Table 4.

TABLE 4

Summary of Rater Background

Rater	Yrs. of Teaching Experience	Nationality
1	2	United States
2	7	Canada
3	4	United States

IV. RESULTS

1. Descriptives

Descriptive statistics for the novice group and the advanced group are summarized in Table 5 and Table 6, respectively. Discourse and pronunciation received the highest scores for the novice group, while overall comprehension and discourse were the highest for the advanced group. The skewness and kurtosis for each variable was examined and there were no values greater than an absolute value of one, suggesting reasonably normal distributions.

TABLE 5

Descriptive Statistics for Novice Group

	N	Minimum	Maximum	Mean	Std. Deviation
Overall Comprehension	30	12	24	19.63	3.14
Grammar	30	12	23.5	18.55	2.72
Pronunciation	30	15	24	20.25	2.37
Fluency	30	11	23	18.97	2.69
Discourse	30	16	25	20.30	2.58
Sociolinguistics	30	13	24	18.87	2.79
Valid N (listwise)	30				

TABLE 6

Descriptive Statistics for Advanced Group

	N	Minimum	Maximum	Mean	Std. Deviation
Overall Comprehension	29	15.5	30	22.88	3.68
Grammar	29	12	29	22.07	3.77
Pronunciation	29	11	30	21.76	4.25
Fluency	29	15	29	21.91	4.07
Discourse	29	13	29	22.48	3.95
Sociolinguistics	29	14	29	21.10	4.08
Valid N (listwise)	29				

2. Inter-rater Reliability

Inter-rater reliability is used to assess the degree to which different raters/observers give consistent estimates of the same phenomenon. The general rule of thumb is that coefficient alpha estimates should always equal or exceed .70 (Cronbach, 1951). An alpha measure lower than .70 generally indicates poor scale reliability. Table 7 shows that a low inter-rater reliability in the assessment of novice learners, while that of advanced learners attained satisfactory reliability.

TABLE 7

Inter-rater Reliability Statistics

	Novice	Advanced
Comprehensibility	.527	.762
Pronunciation	.498	.899
Fluency	.582	.847
Grammar	.521	.814
Discourse	.405	.843
Sociolinguistics	.383	.835

3. Multiple Linear Regression

Multiple linear regression was employed to help determine which of the five elements were the best predictors of overall comprehension. Since no *a priori* hypotheses had been made to determine the order of entry of the predictor variables, a direct method was used for the multiple linear regression analyses. In the novice group, the five elements obtained an adjusted R^2 of .753 (F (5, 24) = 18.666, p = .000) for the prediction of overall comprehension, while the advanced group produced an adjusted R^2 of .933 (F (5, 23) = 78.847, p = .000) for the prediction of overall comprehension.

The best predictor of oral proficiency among novice learners was sociolinguistics (β=.580), followed by fluency (β=.529) (see Table 8).

TABLE 8
Standard Multiple Regression of Oral Proficiency Assessment Criteria on Overall Comprehension (Novice Group)

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	.247	2.759		.089	.929
Grammar	-.113	.185	-.098	-.611	.547
Pronunciation	-.175	.170	-.132	-1.029	.314
Fluency	.617	.195	.529	3.165	.004
Discourse	.051	.181	.042	.283	.779
Sociolinguistics	.651	.178	.580	3.656	.001
Dependent Variable: Overall Comprehension					

Table 9 shows that the strongest predictors of oral proficiency among advanced learners were grammar (β=.322) and fluency (β=.314).

TABLE 9
Standard Multiple Regression of Oral Proficiency Assessment Criteria on Overall Comprehension (Advanced Group)

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	2.013	1.142		1.764	.091
Grammar	.315	.130	.322	2.426	.023
Pronunciation	-.076	.132	-.087	-.571	.574
Fluency	.284	.132	.314	2.154	.042
Discourse	.241	.149	.258	1.615	.120
Sociolinguistics	.186	.116	.207	1.606	.122
Dependent Variable: Overall Comprehension					

V. DISCUSSION

The current study investigated the extent to which assessment criteria for oral proficiency had an influence on overall comprehensibility among novice and advanced learners of English.

The raters who assessed students' oral proficiency showed high inter-rater reliability when assessing advanced learners. However, it was considerably lower in the assessment of novice learners. The best predictor of oral proficiency among novice learners was sociolinguistic competence and fluency, while grammatical competence and fluency were the strongest predictors for advanced learners.

The difference in the outcome of inter-rater reliability draws attention in that it implies the need for different sets of assessment criteria depending on the oral proficiency level of

test-takers. While the high inter-rater reliability of the criteria in the assessment of advanced learners reaffirms its validity for advanced learners, the criteria used in this study might result in inconsistent results for lower level learners.

It is also worth noting that while fluency was an important predictor of oral proficiency scores for both novice and advanced learners, grammar was the decisive factor for advanced learners and sociolinguistic competence was the strongest predictor for novice learners. The results indicate that curriculum developers should take into consideration the different needs of learners as they progress from one level to the other. Unlike conventional thought, accurate grammatical usage in oral communication should be reinforced in the advanced learners' classroom, rather than in the novice level. This indicates that grammatical competence stands at the highest level in difficulty in attaining oral proficiency.

On the other hand, sociolinguistic competence stood out as the strongest element that set learners apart when assessing novice learners. Raters mainly looked for usage of inappropriate or expressions that were out-of-place when assessing sociolinguistic competence and it appears that lack of sociolinguistic competence among novice learners were the biggest obstacles during oral communication.

VI. CONCLUSION

According to ACTFL (American Council on Teaching of Foreign Languages) Guidelines (1999), proficiency is largely regarded as communicative growth. Communicative competence has become the buzzword in the field of English education nowadays and, as a result, recent trends in English education have focused greatly on the communicative competence of English learners. However, the actual assessment of proficiency is anything but a clear-cut matter as with comprehension which is also notoriously hard to define (Iwashita & Grove, 2003).

The question of what determines the comprehensibility of a speech depends on the listener's experience as well as the speaker's communicative ability. As indicated in the results of this study, sociolinguistic competence is a more important factor for lower level speakers than for advanced learners who are more likely to have already possessed a higher level of sociolinguistic competence. The following comments from one of the raters state this point well;

> One oddity I noted was the *constant* use of "grandma" for the white-haired woman in the picture. I'm not sure that's actually accurate: she could just as easily be the mother, and I was thinking maybe it was a transliteration of the more general use of "할머니" (where we might say, "little

> old lady") into English. It seems innocuous here, but in some cases could result in serious miscommunication.
>
> Another case which seems to hinge on cultural difference is the thing where, in lots of cases for the Passport narrative, students claimed that the *husband* had forgotten his passport. (Which was weird to me, since the picture clearly shows the wife, or girlfriend, realizes it's missing and then handles the whole business of retrieving it). But then I remembered my Korean friend pointing out to me in an airport just how many wives carry their husband's passports here.) I wondered if it was just a he/she mix-up, but it seemed pretty consistent. Again, this is a question of possibly cultural barriers to communication. But in any case, it struck me again that what *looks* like relatively clear communication could actually be completely wrong.

If the level of one's oral proficiency shows different needs, i.e., sociolinguistic competence for novice learners and grammatical competence for advance learners, it stands to reason that instructors should place priority in these specific elements when planning for speaking classes. The discrepancy between inter-rater reliability is another issue that should be addressed by the testing agencies/organizations. While the assessment criteria used in this study was a viable assessment tool for advanced learners, it proved less than appropriate for the assessment for novice learners. Either a different set of criteria or a training session to raise inter-rater reliability should be developed to address this issue.

The relatively small number of subjects who participated in the research is one of the limitations of this study. Further research should be conducted on a more extensive scale to include more subjects and more raters of various backgrounds. The medium of the speech, i.e., recordings in an MP3 file, is another compromising factor. Research testing the efficiency of the use of MP3 recordings in EFL context supports the use of such medium (Song, 2008). However, more teachers generally show preference for the Face-To-Face Interview (FTFI) over the Computerized Oral Test (COT), although perceptions of the COT were generally favorable (Joo, 2008). Also, this research should be applied to different forms of oral proficiency, such as, dialogue, expressing opinions, or interviews so that comprehensive assessment criteria for oral proficiency may be completed.

REFERENCES

American Council on the Teaching of Foreign Languages. (1999). *ACTFL proficiency guidelines*. Hasting-on-Hudson, NY: Author.

Adams, M. L. (1980). Five co-occurring factors in speaking proficiency. In J. R. Firth (Ed.), *Measuring spoken language proficiency* (pp. 1-6). Washington, DC: Georgetown University Press.

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

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Brown, A., Iwashita, N., McNamara, T., & O'Hagan, S. (2003). *Getting the balance right: Criteria in the integrated speaking test*. Paper presented at the Annual Language Testing Research Colloquium, Hong Kong Polytech University, Hong Kong.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*(1), 1–47.

Clark, J. L. D. (1988). Validation of a tape-mediated ACTFL/ILR-scale based test of Chinese speaking proficiency. *Language Testing, 5*(2), 197-205.

Conlan, C. J., Bardsley, W. N., & Martinson, S. H. (1994). *A study of intra-rater reliability of assessments of live versus audio-recorded interviews in the IELTS speaking component*. [Unpublished study commissioned by the Inter-national Editing Committee of IELTS].

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297-334.

de Jong, J. H. A. L., & van Ginkel, L. W. (1992) Dimensions in oral foreign language proficiency. In L. Verhoeven & J. H. A. L. de Jong (Eds.), *The construct of language proficiency* (pp. 187-205). Amsterdam: John Benjamins.

Douglas, D. (1994). Quantity and quality in speaking test performance. *Language Testing, 11*(2), 125–144.

Higgs, T. V., & Clifford, R. (1982). The push toward communication. In T. V. Higgs (Ed.), *Curriculum, competence, and the foreign language teacher* (pp. 57-79). Lincolnwood, IL: National Textbook.

Iwashita, N., & Grove, E. (2003). A comparison of analytic and holistic scales in the context of a specific purpose speaking test. *Prospect, 18*(3), 25-35.

Joo, M. (2008). The need for an alternative approach to oral testing. *English Language & Literature Teaching, 14*(1), 1-20.

Kenyon, D., & Malabonga, V. (2001). Comparing examinee attitudes toward computer-assisted and other oral proficiency assessments. *Language Learning & Technology,*

5(2), 60-83.

Kenyon, D., & Tschirner, E. (2000). The rating of direct and semi-direct oral proficiency interviews: Comparing performance at lower proficiency levels. *Modern Language Journal, 84*, 85-101.

Kim, H. (2009). The role of grammar in speaking proficiency: Exploring the need for spoken grammar. *Modern English Education, 10*(1), 46-65.

McNamara, T. (1990). Item response theory and the validation of an ESP test for health professionals. *Language Testing, 7*(1), 52–75.

Song, J. (2008). Student-centered speaking practice using an MP3 player and the text-bulletin board. *English Language & Literature Teaching*, *14*(1), 75-95.

Thompson, I., Thompson, R.T., & Hiple, D. (1988). Issues concerning the less commonly taught languages. In P. Lowe, Jr., & C. W. Stansfield (Eds.), *Second language proficiency assessment: Current issues* (pp. 83-124). Englewood Cliffs, NJ: Prentice Hall Regents/Center for Applied Linguistics.

Wylie, E. (1993). *Study of the inter-rater reliability of the IELTS speaking test*. [Unpublished study commissioned by the International Editing Committee of IELTS].

APPENDIX A

Assessment Grid for Speaking Proficiency

Pronunciation

Grade	Description
	Consonants/Vowels & Pronunciation/stress/intonation
9~10	Pronunciation and intonation approximate those of a native speaker.
7~8	Accent characteristic of non-native speakers is sometimes detected. However, the content is 100% comprehensible.
5~6	Pronunciation errors and stress/intonation flaws are detected. The content is fairly understandable.
3~4	Frequent pronunciation errors are detected. Stress/intonation flaws often impeded communication.
1~2	Due to pronunciation/stress/intonation errors, the message is incomprehensible.

Fluency

Grade	Description
	Strategic use of language, natural flow of speech, use of sophisticated expressions
9~10	Natural flow of speech approximates that of a native speaker.
7~8	Speaks naturally uses good expressions with few lapses.

5~6	Shows occasional lapses in speech but does not impeded communication.
3~4	Stammers, uses unconventional expressions which impede communication.
1~2	Frequent lapses, lack of connection, unnatural flow of speech make it all but impossible to understand the speaker.

Grammatical Competence

Grade	Description
	Practical use of grammar: Accuracy of phrases and grammatical usage
9~10	Rarely shows any grammatical errors in speech.
7~8	Makes occasional grammatical errors. However, such errors are also often exhibited by native speakers.
5~6	Makes grammatical errors. However, the errors are not serious enough to impeded communication. Overall, the structure of speech is acceptable.
3~4	The speaker has basic knowledge of grammar, but makes serious errors that impede communication.
1~2	Except for simple phrases, the speaker fails to make any grammatically correct sentences.

Discourse Competence

Grade	Description
	Logical flow of discourse
9~10	Can use conjunctions in a natural manner so that the message comes across logically.
7~8	Displays fairly competent use of conjunctions and maintains consistency in conveying the message.
5~6	Can use simple forms of conjunctions and can more or less convey the message.
3~4	Fails to use appropriate forms of conjunctions, which makes the message choppy.
1~2	Uses inappropriate forms of conjunctions or fails to use any conjunction. The message lacks consistency.

Sociolinguistic Competence

Grade	Description
	Good choice of words and phrases
9~10	Inappropriate or expressions out-of-place are rarely detectable.
7~8	Inappropriate or expressions out-of-place are sometimes detectable.
5~6	Makes errors in the choice of words or formality in speech.
3~4	Makes many errors in the choice of words or formality in speech.
1~2	The majority of the speech uses inappropriate words and expressions.

Overall Comprehensibility

Grade	Description
	Comprehensive Assessment
9~10	Shows few errors in the use of language and pronunciation. The message comes across perfectly.
7~8	Shows grammatical/pronunciation errors time to time, but the message is easily understandable.
5~6	Displays pronunciation, grammar, vocabulary errors. Shows stops and backtracking. However, overall message is communicated.
3~4	Frequent stops and backtracking. Pronunciation errors, poor choice of vocabulary, inefficient use of grammar. Overall communication is hampered.
1~2	Extremely low level of communication. Cannot provide appropriate answer to the question.

Source: TOP (Test of Oral Proficiency in English) administered by the Seoul National University

Examples in: English
Applicable Languages: English
Applicable Levels: Secondary/Post-secondary

Haeyoung Kim
The Catholic University of Korea
Dept. of English Language and Culture
Bucheon City, Gyeonggi-do
Tel: (02) 2164-4824
Email: haeyoungkim@catholic.ac.kr

Received in April, 2009
Reviewed in May, 2009
Revised version received in June, 2009