

# RELIABILITY TEST ON SPEAKING PROFICIENCY FOR ENGINEERING STUDENTS OF POLINES

Sugeng Irianto

A lecturer at Mechanical Engineering Department, Semarang State Polytechnic  
Jl Prof H Sudharto, S.H., Tembalang Semarang Telepon (024)-7473417, 7466420  
(hunting), Fax (024)-7472396 <http://www.polines.ac.id>; E-mail:secretariat@polines.ac.id

## Abstract

English language skills (listening, speaking, reading and writing), according to the result of survey on stake holders about the need analysis of company towards active ability of English for the graduates of State Polytechnic of Semarang conducted by UPT Bahasa (Language Centre) in 2004, must be taught and tested. For the speaking test, UPT Bahasa uses the analytic approach that is the one which considers that any aspect attached to the English speaking ability basically develops at different rate so it should be evaluated separately. This research tries to uncover whether the approach has higher reliability and efficiency than the other approach i.e. holistic approach which considers speaking proficiency as a general performance and that it is represented by a single score. The data used are taken from the result of speaking assessment of 60 participants from the 2<sup>nd</sup> and 3<sup>rd</sup> grade of Civil, Electronic and Mechanical Engineering. Based on the level of reliability coefficient 0.70 – 0.79 (Hughes, 1989), the result of the research shows the analytic approach has a correlation mark of 0.803 at level of significance 0.000, whereas holistic approach points at 0.653 correlation at the significant level of 0.000 which means that the analytic approach is reliable; but the holistic is not since the correlation points at less than 0.70.

**Keywords:** speaking, test, reliability

## Introduction

According to the result of survey on stake holders about the need analysis of company towards active ability of English for the graduates of State Polytechnic of Semarang conducted by UPT Bahasa (Language Centre) in 2004, English language skills (listening, speaking, reading and writing) must be taught and tested. But in reality, English lecturers at Semarang State Polytechnic (Polines) and at some other education institution are reluctant to conduct the assessment of productive skills of English such as *speaking* and *writing* because this type of assessment will require much time, energy, and money.

For the purpose, at the end of 2004 UPT Bahasa Polines constructed the *English Proficiency Test (EPT)* for *Engineering Students of Polytechnic* of semester 4 or above which covered the four English language skills (speaking, reading, writing, reading). This EPT is used to get the level of mastery of productive English among students, especially students who are going to graduate from Polines.

So far, for maintaining the high level of reliability of the speaking test, UPT Bahasa uses the analytic approach that is the one which considers that any aspect attached to the English speaking ability basically develops at different rate so it should be

evaluated separately. This kind of approach needs more time, energy, and cost than the holistic approach, i.e. an approach which assesses the speaking ability not in sections or items of mastery but in general in the form of one item of the mark (one item evaluation). But in the process of the dissemination of the analytic approach, there were many complaint from graders (lecturers) about the difficulties to apply the approach. To get the belief of which appropriate approach should be conducted for *speaking test* i.e. the system of speaking test with high reliability and efficient, the research was conducted in order to test the reliability of the marking system of students' speaking ability using holistic versus analytic approaches.

### **Speaking Test**

The result of stakeholders survey shows that the mastery of speaking should be gained by Polines alumni. So, the teaching of English in this institution must cover the mastery of speaking. Furthermore, the assessment conducted must cover the four language skills i.e. speaking, reading, listening, and writing. To get the picture of the ability of the student in mastering English language, a *test battery* which contains *test-structure*, *extended writing*, and *listening test* and also *oral*

*test/speaking test* is used (Underhill, 1988). There are many format for testing the speaking ability of students, some of them are Verbal Essay, Oral Presentation/report, The Free Interview, The Controlled Interview, Information Transfer, Description of a Picture Sequence, Information Transfer, Question on a Single Picture, Interaction Task, and Role Play (Cyril J. Weir, 1990).

*Marking system* is the most important part of oral test. The major problem of Test Marking on Speaking Test is the doubtful from many parties especially test takers towards the consistency of the graders in doing their jobs. The test of language ability with the highest subjectivity is the oral test and extended writing. Two graders may give different mark on the same test taker. The low level of reliability causes the test takers is difficult to believe that the mark given by the graders will be appropriate.

The solution of the problem is by doing a good and appropriate *Marking System*. This marking system should be made in good consistency with the goal of the system. It should also appropriate with the kind of oral test conducted by the test takers. In designing Marking System some factors should be matched with the need of test takers, the funding availability, and the

reliability of the standard. Underhill (1988) mentions some factors which influence the reliability of the Marking System such as *the numbers of assessor/graders* (using more than one graders will increase the reliability), *the choosing and skill giving of the graders*, *assessing category*, and *weighing or Rating Scales*.

## The Marking Approach

There are two kinds of marking approaches i.e. holistic and analytic system. The following are the review on the differences of the two approaches.

Table 1.  
Scale Comparison of Holistic and Analytic

<i>Quality</i>	<i>Holistic Scales</i>	<i>Analytic Scales</i>
<i>Reliability</i>	<i>lower than analytic, but still acceptable</i>	<i>higher than holistic</i>
<i>Construct Validity</i>	<i>assume that all relevant aspects of speaking ability develop at the same rate and can thus be captured in a single score;</i>	<i>more appropriate for L2 speakers as different aspects of speaking ability develop at different rates</i>
<i>Practicality</i>	<i>relatively fast and easy</i>	<i>time-consuming; expensive</i>
<i>Impact</i>	<i>single score may mask an uneven speaking profile and may lead to misleading placements</i>	<i>more scales provide useful diagnostic information for placement and/or instruction; more useful for rater training</i>
<i>Authenticity</i>	<i>White(1995) argues that listening holistically is a more natural process than listening analytically</i>	<i>Raters may listen holistically and adjust analytic scores to match holistic impressions</i>
<i>Interactiveness</i>	<i>N/a</i>	<i>n/a</i>

Source: adapted from Weigle 2002, p.121

For the reason of cost and time, it is suggested to use the holistic (one item evaluation) approach but to avoid the risk of the assessment (idiosyncratic) the analytic approach (with several evaluation items) is suggested. In the practice of speaking test using holistic approach, the mark is given in the form of a number or a letter. Whereas analytic approach gives

mark which contain some numbers represents the aspects of the marking through the speaking task which decided before (Hughes, 1989). Until now, there are some doubtful in using the appropriate approach (analytic or holistic) in testing speaking ability of the test takers. The two approaches have their own advantages and disadvantages in testing of speaking.

Holistic approach is practically done because the graders do not need complex thinking to give the score/grade or mark to every single test taker and also do not take much energy and time for conducting this approach for testing speaking. Contrary to holistic approach, analytic approach needs more thinking, energy, and time. This approach is chosen because of some advantages such as to ability to cover the level of subskills of every single test taker. The analytic approach is also regarded to be more reliable because graders should give the mark for every aspect or mark items which decided before. Also, the analytic approach is believed to be free from "hallo effect" (Hughes, 1989).

### Research Methodology

The research or study uses 60 students of Engineering (Civil, Electric, and Mechanical) Semarang State Polytechnic who took Speaking Test. There are four score lists as research variables i.e. 2 speaking test scores taken from the results of assessment (marking) using holistic approach given by Grader A and Grader B, and 2 other speaking test scores taken from the results of assessment using analytic approach given by Grader C and Grader D. Making scheme of the speaking test is regarded to be reliable if it has the level of

reliability from 0.70 until 0.79 (Hughes, 1989). The test of reliability system of speaking test is as follows:

- a. Two graders (Grader A and Grader B) separately give the mark using holistic approach. The mark from Grader A and Grader B are compared to get the level of reliability. The higher the level of reliability coefficient, the less the difference of the marking given by Grader A and Grader B (pair correlation method).
- b. Two graders (Grader C and Grader D) separately give the mark using analytic approach. The mark from Grader C and Grader D are compared to get the level of reliability. The higher the level of reliability coefficient, the less the difference of the marking given by Grader C and Grader D (pair correlation method).
- c. Comparing the level of reliability coefficient *point a* and *point b* : the approach having the higher reliability coefficient is regarded to be more reliable.

### Marking Scheme

The marking scheme using holistic approach used in the study can be seen in the following table.

Table 2.

## Marking Scale using Holistic Approach

Level 1	<i>Very limited personal conversation. Knows formulaic greetings and some vocabulary. Cannot construct correct simple sentences.</i>
Level 2	<i>Personal and limited social conversation. Can answer simple questions about personal topics correctly in present and past tenses. Has difficulty with question formation. Vocabulary still very limited.</i>
Level 3	<i>Basic competence for social and travel uses. Has basic command of all simple tenses and can operate question and negative forms. Shows awareness of perfect forms but makes errors in using them. Familiar with common concrete vocabulary: still searches for words.</i>
Level 4	<i>Elementary professional competence. Makes effective use of all tenses, including past vs perfect and simple vs continuous distinctions; occasional errors in tense forms.</i>
Level 5	<i>General proficiency on all familiar and common topics; may be at a loss for words on other topics, but is able to paraphrase successfully. Can produce correct complex sentences; very rare errors in structural forms, but makes errors of idiom or collocation.</i>

Sources: Nic Underhill, 1988

Marking scheme using analytic approach used in this study is shown in the table 3 below. There are 7 component of speaking ability being marked namely *task, coherence, strategy, pronunciation, grammar, vocabulary* and *fluency*.

Table 3.

## Marking Scale using Analytic Approach

No	Komponen	Level										
		1	2	3	4	5	6	7	8	9	10	
1	Task											
2	Coherence											
3	Strategy											
4	Pronunciation											
5	Grammar											
6	Vocabulary											
7	Fluency											

**Discussion**

Holistic approach is shown in the following table 4.

## The Test Result

The result of the speaking test for 60 respondents (test takers) using Analytic and

**Table 4**  
**The Result of Speaking Test Using**  
**Analytic and Holistic**

Respondent's number	A1	A2	H1	H2
1	6.6	3.5	5	8
2	6.9	5.3	6	8
3	3.1	2.6	5	6
4	3.5	2.8	4	5
5	5.9	2.9	5	6
6	1.0	1.2	1	1
7	2.0	2.4	3	5
8	5.9	2.8	3	5
9	3.7	2.5	3	5
10	5.8	4.0	6	6
11	1.9	1.8	4	4
12	2.1	2.0	3	5
13	3.0	2.6	3	5
14	5.5	2.8	6	6
15	5.4	3.5	5	5
16	4.3	3.5	6	7
17	7.7	4.6	6	6
18	3.7	3.1	5	5
19	5.9	2.6	6	5
20	4.5	3.1	4	5
21	5.3	3.3	5	7
22	5.2	3.3	6	6
23	6.4	4.5	6	8
24	3.0	2.7	6	5
25	4.2	2.7	5	5
26	6.1	2.5	6	6
27	5.4	4.6	5	7
28	4.0	2.5	3	4
29	5.1	3.1	6	5
30	1.9	1.7	3	4
31	4.6	2.7	4	6
32	8.0	5.6	7	10
33	8.0	6.2	8	9
34	5.8	3.0	5	5
35	4.0	4.0	6	7
36	6.6	2.7	6	4
37	6.2	3.9	5	6
38	5.8	4.3	6	9
39	4.5	2.8	5	6
40	4.9	2.8	3	3

41	8.3	4.6	9	7
42	8.4	5.2	7	8
43	3.1	2.6	4	3
44	5.5	3.8	7	7
45	2.4	2.3	3	4
46	3.6	2.5	4	7
47	3.9	2.8	5	5
48	5.4	3.0	6	6
49	3.0	2.7	5	6
50	2.0	2.3	3	4
51	5.6	4.5	5	7
52	6.2	3.9	6	7
53	2.0	2.0	4	7
54	5.5	2.9	6	6
55	5.0	2.5	6	6
56	5.2	2.9	3	5
57	6.1	4.6	7	5
58	3.4	2.9	3	5
59	6.3	4.3	7	7
60	5.3	2.9	6	5

The score of A1 and A2 are the analytic test result done by Grader C and Grade D, whereas the scores of H1 and H2 are the holistic result done by Grader A and Grader B.

#### **Reliability Test of Holistic and Analytic Approach**

The following are the tables of the data analysis using the instrument of SPSS11 Program. Table 5, Table 6, and Table 7 are T-Test Tables, while Table 8, Table 9, Table 10, Table 11, Table 12, and Table 13 are Correlation Tables, and Table 14, Table 15, Table 16, and Table 17 are Reliability Tables.

Table 5

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 A1	4.827	60	1.7544	.2265
A2	3.228	60	1.0058	.1298
Pair 2 H1	5.02	60	1.513	.195
H2	5.77	60	1.566	.202

Table 6

Paired Samples Correlations

	N	Correlation	Sig.
Pair 1 A1 & A2	60	.803	.000
Pair 2 H1 & H2	60	.653	.000

Table 7

Paired Samples Test

	Paired Differences					t	df	Sig (2-tailed)
	Mean	Std Deviation	Std. Error Mean	95 % Confidence Interval of the Difference				
				Lower	Upper			
Pair 1 A1-A2	1.5983	1.1208	.1447	1.3088	1.8879	11.047	59	.000
Pair 2 H1-H2	-.750	1.284	.166	-1.082	-.418	-4.525	59	.000

Table 8  
Correlations

	A1	A2
A1 Pearson Correlation	1	.803**
Sig. (2-tailed)	60	60
N		
A2 Pearson Correlation	.803**	1
Sig. (2-tailed)	.000	60
N	60	

\*\* Correlation is significant at the 0.01 level

Table 10

Test for Model Goodness of Fit

Chi-Square	Value	38.833
	df	1
	Sig	.000
Log of Determinant Matrix of	Unconstrained	.101
	Constrained Matrix	.777

Under the parallel model assumption

Table 11

Reliability Statistics

Common Variance	2.045
True Variance	1.417
Error Variance	.628
Common Inter-Item Correlation	.693
Reliability of Scale	.819
Reliability of Scale (Unbiased)	.825

The covariance matrix is calculated and used in the analysis

Table 9  
Correlations

	H1	H2
H1 Pearson Correlation	1	.653**
Sig. (2-tailed)	60	60
N		
H2 Pearson Correlation	.653**	1
Sig. (2-tailed)	.000	60
N	60	

\*\* Correlation is significant at the 0.01 level

Table 12  
Inter-Item Correlation Matrix

	A1	A2
A1	1.000	.803
A2	.803	1.000

The covariance matrix is calculated and used in the analysis

Table 13  
Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Inter-Item-Correlations	.803	.803	.803	.000	1.000	.000	2

The covariance is calculated and used in the analysis.

Table 14  
Reliability  
Case Processing Summary

		N	%
Cases	Valid	60	100.0
	Excluded <sup>a</sup>	0	.0
	Total	60	100.0

a. Listwise deletion based on all variables in the procedure.

Table 15  
Reliability Statistics

Cronbach's Alpha	Part 1	Value	.819
		N of Items	2 <sup>a</sup>
	Part 2	Value	.790
		N of Items	2 <sup>b</sup>
Total N of Items			4
Correlation Between Forms			.813
Spearman-Brown Coefficient	Equal Length		.897
	Unequal Length		.897
Guttman Split-Half Coefficient			.896

a. The items are: A1, A2

b. The items are: H1, H2

Table 16

## Inter-Item Correlation Matrix

	A1	A2	H1	H2
A1	1.000	.803	.757	.619
A2	.803	1.000	.715	.757
H1	.757	.715	1.000	.653
H2	.619	.757	.653	1.000

The covariance matrix is calculated and used in the analysis

Table 17

## Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum/ Minimum	Variance	N of Items
Inter-Item Part 1	.803	.803	.803	.000	1.000	.000	2 <sup>a</sup>
Correlation Part 2	.653	.653	.653	.000	1.000	.000	2 <sup>b</sup>
Both Parts	.717	.619	.803	.184	1.298	.004	4

The covariance matrix is calculated and used in the analysis

c. The items are: A1, A2

d. The items are: H1, H2

### Conclusion

From the Summary Item Statistic using SPSS11 Program above, it shows that the point of correlation assessment using Analytic Approach (Part 1) is 0.803 at level of significance 0.0000, whereas Holistic Assessment Approach points at 0.653 correlation at the significant level of 0.000. Based on the reliability standard for the speaking test is between 0.70 until 0.79 (Hughes, 1989) so the final conclusion approves that the Analytic Approach is

reliable (above the reliability standard), but the Holistic Approach is not since the correlation points at less than 0.70.

Based on the conclusion, it is suggested to use the Analytic Approach in testing the Speaking Ability of the students. But the disadvantages of the approach which probably arise such as the complexity for applying the approach should be eliminated. In the future it is needed to solve the problem in order to make easier for making the concept of sub-skill criteria in the form of checklist for the graders.

## References

- Hughes, A, 1989, *Testing for Language Teachers*, Cambridge: Cambridge University Press.
- Nur Rini, dkk. 2001, *Kajian Penggunaan Bahasa Inggris Bisnis Jurusan Administrasi Niaga Politeknik Negeri Semarang di Dunia Kerja*, (Laporan Penelitian).
- Underhill, N. 1988. *Testing Spoken Language*. Cambridge: Cambridge University Press.
- UPT Bahasa, 2004, *Survei Stake Holders dalam Rangka Analisis Kebutuhan terhadap Kemampuan Bahasa Inggris Lulusan Politenik Negeri Semarang* (Laporan Penelitian).
- Weir, J.C., 1990, *Communicative Language Testing*. London: Prentice Hall.