Designing a Marking Scheme to Assess the Performance of TE Basic Students at ILP

Alvin Taufik

Bunda Mulia University

Abstract

The aim of this project work is to design a marking scheme that will be used in assessing the performance of Talking English (*TE*) Basic students in ILP. The scheme was designed using the stages proposed by Mertler (2001). The stages start with the process of deciding the aim of marking scheme creation in order to acknowledge the importance of the proposed scheme. It is then continued with the process of deciding the object of the assessment, the type, criteria, scale and performance description of the scheme. The steps end with the testing of the scheme. The tested scheme is perfected through discussion of the results of the testing. The criteria of the scheme are formulated based on the concept of Communicative Language Ability as proposed by Bachman (1990). Based on the formulation, the writer has created a descriptive based marking scheme which distribution is done daily through the use of check list method.

Keywords

Marking Scheme, Communicative Language Ability, Performance

1. Introduction

Talking English (*TE*) is one of the classes in ILP which main purpose is to create unique and daily linguistic situation and to increase fluency of the students. As its name suggests, the class is designed to provide a situation in which the students can perform their skills in speaking English. In TE classes students are demanded to be proactive in learning. Unlike general classes which demands continuous assessment, in TE class, for the students to pass a level, they have to be active. In other word, there are no summative assessments in TE classes.

TE classes are basically designed to facilitate the practice of students' language ability in GE classes. This does not mean that students who have learnt English in other institution cannot join TE classes. For students who have acquired training in other institution, can join the class if they meet the requirements. One of

the requirements is to have passed the entrance test which is equal to the entry test required for GE Basic 2 class in ILP.

TE classes are required to have 10 meetings in a single term given within duration of 120 minutes per meeting. There are 6 levels of TE classes; the first and second is named Basic classes. The third and fourth classes are called Intermediate classes, and the fifth and last are called Advance classes. There are no apparent reasons for the division of the classes

The materials given in TE classes, for all classes, are divided into four main activities; they are Short Activities, Task-Based Activities, Topic-Based Activities, and Language Awareness Activities. In Short Activities, students are given assignments which can be finished in a short moment, such as discussing vocabulary that will be used for that day's lesson. The main purpose of these Short Activities is to provide useful activities that can be performed in group or individually by the students, and they are usually given at the beginning of the lesson. In Task-Based Activities, students are given tasks which encourage the students, in group or individually, to use English in a situation, such as ordering food from a menu. In Topic-Based Activities, students are given a topic which must be handled using the linguistic competence that they have. As an example, students are put in the restaurant, and they were asked to demonstrate what they can do in that restaurant and what language function can be used to handle the situation. Moreover, in Topic-Based Activities, students can also discuss the topic given by the teachers. At the end of the lesson, the students are given some activities to comprehend language. For the last activities, students are given the chance to review the activities that they have done and learn from the mistake that they have done from the activities. On the last meeting, students are given reports which explain their development in learning English in the classroom.

Eventhough there were no summative assessment, teachers must still have to give some assessment to the students. The assessment given in TE class is in the form of an analytical scheme with analytical criteria. This type of assessment, as mentioned by Allen (2003), Moskal (2003), and Mertler (2001), separate the observable performance components. In addition, the scoring system given in this class is divided into 8 categories, and they are given at the end of the term. According to Underhill (1987), this type of assessment is summative assessment. Looking at that fact, it can be said that the assessment given in TE class is in full contradiction to the function of

analytical scheme proposed by Mertler. That is because a summative assessment should be given in a form of holistical scheme. Moreover, the division of the category does not have detailed explanation. As a consequence, when the teachers need to input the score into the scheme, their judgment will be very subjective. Another existing problem is that the scoring will not be significant if related to the passing criteria of the class. This is because the scoring will not affect the passing requirement of the class.

Another questionable point concerns with the amount of criteria in the current scheme. As previously mentioned, there are 8 criteria used in the current scheme. Underhill stated that a marking scheme should not contain more than 4 criteria. Moreover, Heaton (1975) also mentioned that an effective scoring scheme is the one that does not require the teachers to look at the scheme too often.

Another difficult task for the teachers is to convert the scheme into numbers. Mertler mentioned that the conversion is a creative process. Therefore, teachers have to be creative in setting the conversion system to be used. The conversion system has to be adapted to the need and aim of the lesson.

In addition, Moskal (2000) also mentioned that designing a scheme can also be based on the purpose of giving the lesson itself. This is because the students need to know the progress that they have made in following a lesson. Therefore, the score must also function as a feedback for the students. So, the number in the score must have a clear definition.

Based on those facts, it can be seen that the current marking scheme implemented in TE classes in ILP has not fully helped the teachers in giving quantitative scoring to the students. The scoring standard is not available in the institution so the teachers must often rely on his/her experience in giving the score. Based on that, the researcher wishes to design a marking scheme which has a clear standard and can be used in giving assessment that can be used daily in Talking English class.

2. Formulation of Research

The main problem of the current marking scheme is the insignificancy that it poses due to the non-influential nature of the scheme. The other problem is that the number in the scheme which has been used to score daily performance of TE students has no qualified criteria. Therefore, the TE classes in ILP in general need scoring system

which has detailed criteria so that the feedback presented by the scheme will be beneficial for the students and teachers alike. In addition, the distribution of the assessment needs to be changed from summative into on-going assessment.

Based on the existing problem, the researcher has formulated several statements of problem, they are

1. What to score and what kind of assessment should be administered to the students on daily basis

2. What sort of criteria should be included in the scheme which will be more comprehensive from the teachers and could represent the scoring of students' performance.

3. Research Objective

This research is aimed to give contribution to the scoring system in TE classes in ILP. Moreover, the research will hopefully provide access for teachers of conversation classes in giving a more objective assessment. It is expected that proper implementation of the scheme will provide higher reliability in assessing students' performance.

4. Scope of Research

The research will focused on the design of a marking scheme that can be used daily in TE classes in ILP. There are 3 levels in TE classes; Basic, Intermediate, and Advance. The estimations of performance of the three classes are different. The scheme used in a level cannot be used on other levels. Based on that thought, the research will be limited on the design of marking scheme for TE Basic classes.

The data analyzed will be taken from one TE Basic class. The data will later be used to test the scheme and the result of the scheme testing will be discussed in the conclusion of this research.

5. Research Method

As previously mentioned, the aim of the research is to design a marking scheme to assess students' performance in TE classes in ILP. There are two kinds of data collected to support the research. The primary data is used in designing the scheme, while the secondary is used to test the scheme. The primary data is collected in the

form of recording of the students' performance and interview with the teachers, as well as the recapitulation of the questionnaires.

In later paragraphs, there will be simplified description of data collected. Those descriptions contain some explanation on the respondents, research instruments and data processing.

6. Respondents

This research involves teachers who have taught in TE classes previously. The total respondents are 5 individuals. The testing of the scheme was participated by 4 female respondents and 1 male respondent. Figure 1 shows the comparative descriptions of the amount of respondents based on sex.
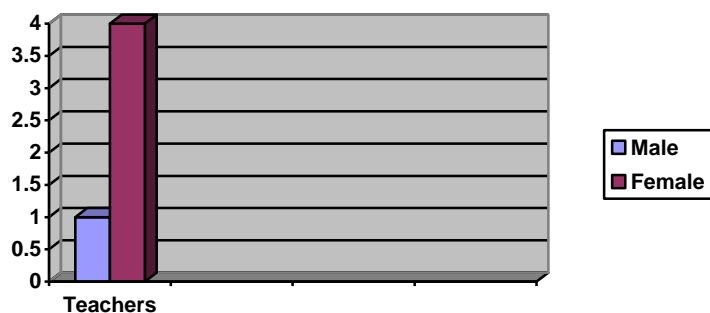

Figure 6.1 comparative description of respondents based on sex

Moreover, in the testing of the scheme, the teachers are demanded to have some knowledge on the concept of competence so that they will be able to use the scheme appropriately. Based on the fact, the teachers are given a questionnaire which basically tests their knowledge on such concept. Figure 2 shows the comparative description of the teachers' knowledge on the concept of competence.
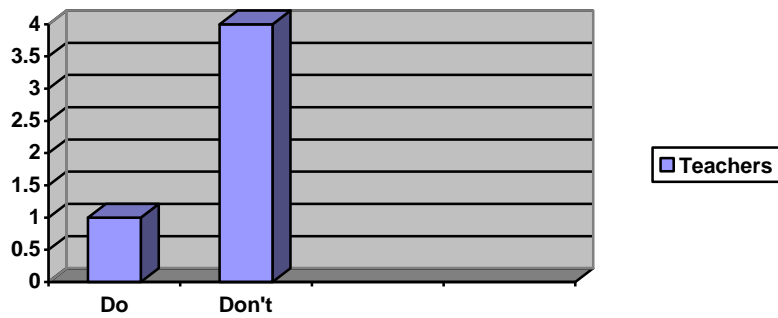
Figure 6.2 comparative description of the teachers' knowledge on competence.

7. Research Instrument

The primary research instrument in this research is the questionnaire and the video camera recorder. The questionnaire was divided into 4 sections in which each section it includes the introduction of the researcher, the respondents' personal data, and detailed questions on competence and essay questions on the comprehensiveness of the concept of competence. The Video camera recorder was used mainly in gathering the primary data required for designing the scheme.

8. The content of the questionnaire

The questionnaire contains 18 questions. There are 5 questions for the second part of the questionnaire, 12 questions on the third part and 1 question on the last part. Part 1 contains various information of the respondents, such as educational background, sex, name, age, and TE classes that they have taught. The questions on part 2 concern mainly on the competence which become the focus of research.

9. Data Processing

There are primary and secondary data used in this research. The primary data was acquired through the use of questionnaire. The data was then processed to become the basis of scheme formulation. The result of this primary data processing will be used as secondary data.

Another data needed to determine whether the scheme is usable or not is the statistical data. The statistical data is classified from secondary data and is acquired through the testing of the scheme. The statistical data was later used to determine the

reliability of the scheme. Processing the secondary data into statistical data was done through several phases.

The data processing phase starts with the collection of secondary data. The collected data was then sorted and any incomplete data was removed. There are 5 data collected from the scheme testing. The calculation of the data was done using computer. Data collected from the scheme testing was qualitative. In order to account for the reliability of the scheme, the data needs to be made quantitative. The quantitative data was acquired using the formula proposed by Salkind (2004).
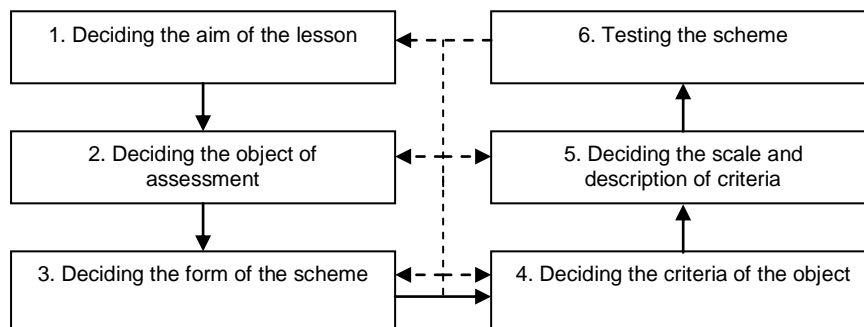
10. The Method in Designing Marking Scheme

In this research, the marking scheme was designed based on the observation made from the recording of students' performance and the interpretation of primary data. The marking scheme was also designed in regard to the principle of practicality, reliability and validity. The formulation of marking scheme based on the expected students' performance was proposed by both Moskal (2000) and Mertler (2001). Therefore, the steps in designing the scheme in this research were also based on the steps of marking scheme formulation as stated by Mertler.

Mertler stated that the process of scheme designing starts with determining the aim of the lesson in TE Basic classes. After the aim has been determined, the next stage is to determine the criteria to be assessed. According to Moskal and Brown (1994), there are several items that can be assessed in the scheme, and one of them is the performance of student.

The next step is to determine the object of assessment and the amount of performance criteria which should be included in the scheme. As previously mentioned, the performance criteria included will be based on the expected performance of the students joining the class. In between those process is the process of determining the form of the scheme. Two common types of scheme used are holistical and analytical scheme. The next step is to determine the scale and description of the inclusive criteria. The last of the proposed step of designing the scheme is to test the finished scheme to the teachers who are going to use them. Figure 3 shows the steps in designing the scheme based on Mertler's proposition.

Figure 10.1 steps in designing the scheme (Mertler, 2000)

1. Deciding the aim of the lesson	6. Testing the scheme
2. Deciding the object of assessment	5. Deciding the scale and description of criteria
3. Deciding the form of the scheme	4. Deciding the criteria of the object

The dashed line in the diagram shows that there will be some adjustment made to the scheme even after it is finished.

11. The Steps

11.1 Deciding the Aim of the Lesson

In this research the purpose of having the scheme is formulated by observing the learning goal of TE Basic classes in ILP. The data is gathered through interview with the officer in charge of TE program in ILP and through the TE syllabus.

Through the interview and syllabus, it was found that the main purpose of the lesson is to increase students' fluency in speaking English in real-time situation. In addition, the syllabus also mentioned that the purpose of daily activities in class is NOT to teach the right grammar but to increase the naturalness of their English and to give the students the freedom to express themselves in finishing the tasks given by the teachers.

Based on the fact it can be concluded that TE program is focused on the students' performance. Therefore, the scheme should be made to assess performance. Moreover, performance cannot be judged or assessed on one sitting. As a result, the assessment should be done on every meeting.

11.2 Deciding the Object of Assessment

The object of assessment is not only about what to assess but also who will be assessed. This is in relation to the statements in the previous part of this research which explain that the scheme will also be based on the expected performance of the students. The teaching in TE Basic classes is focused on the teaching of expression and basic grammatical structures, such as how to greet or how to express likes and

dislikes. The requirement of the program clearly stated that those who can follow this class are students who have at least passed Basic 2 level GE in ILP. Moreover, according to the interview made with the one of the principal of ILP branch, students who join the class have been tested at the beginning of the lesson. This is to ensure their activity in the classroom.

Based on the fact it can be concluded that the students' proficiency in the classroom can be classified into upper basic level. As a consequence, the designing of the scheme must not include definition of criteria which surpasses that level.

Having formulated the "who" in the scheme, the next step is to determine the "what" in the scheme. One of the deciding factors in designing the scheme is the expected students' performance. Performance is determined by the competence of each individual. The concept of competence in this research refers to the concept proposed by Bachman (1990). They five component proposed by him will be the base of area of performance in the scheme. On table 1, it can be seen the area of performance which will be used in the scheme.

Area of Performance				
Grammatical Competence	Textual Competence	Illocutionary Cmpetence	Sociolinguistic Competence	Strategic Competence

Table 11.2.1 deciding the object of assessment

11.3 Deciding the Form of the Scheme

The process of determining the form of the scheme is influenced by the objects of assessment and the frequency of assessment. In this scheme, the assessment will include the categories of competence of a performance. Moreover, the scheme will also include the scale of competence.

The frequency of assessment is also one of the considerations in designing the scheme. This is related to the choosing of the amount of criteria which should be included in the scheme. Since the scheme is going to be used daily, the scheme should be practical to use and easy to memorize so that everytime the teacher has to assess his or her students, they can do it without having to look at the scheme.

From the objects of assessment that has been decided previously, it can also be concluded that there will be detailed criteria of performance which will be assessed in

the scheme. Therefore, it is best to choose the analytical scheme to assess the students' performance.

11.4 Deciding the Scoring Criteria

The scoring criteria and narrative description in this research is formulated by adapting to the concept of competence proposed by Bachman, while the detail of the concept is arranged based on the explanation of Bachman concept of competence as proposed by Qin and Li (2008)

The first criterion of competence is the grammatical competence. In this research, the concept is renamed into grammatical knowledge. According to Bachman and Palmer (1996) grammatical knowledge concerns with the students' ability in producing and understanding utterance which has proper grammatical structures, vocabulary knowledge and phonological and syntactical comprehension.

Based on the observation, the students' performance is still focused on the teaching of pronunciation, word structures and the mastering of vocabulary. Therefore, the criteria which will be used in the scheme under the heading of grammatical competence

The other criterion is based on the concept of textual competence. In general, textual competence can be interpreted as a person's ability in producing and understanding a series of information in a discourse. Moreover, Qin and Li stated that an individual use of cohesive devices and arrangement of utterance reflect a person's textual competence. Based on that, the researcher put those areas of performance under the heading of textual competence. In addition, the heading will be named Organization.

Another criterion of performance according to Bachman is the illocutionary competence. He stated that illocutionary competence shows someone's ability in interpreting a relationship between an utterance and the use of a language. Basically, this sort of competence describes a person's ability in using the language. In regard to that fact, the name of heading for the competence is changed into language function.

Illocutionary competence is actually a part of a pragmatic competence. The other component of pragmatic competence is sociolinguistic competence. Sociolinguistic competence is interpreted as the ability in producing and interpreting the accuracy in using language in certain setting. One aspect that can be included in this categorization of competence, and one which is related to the requirements in the

syllabus is the naturality of a person in interacting in a society. For that matter, the heading of sociolinguistic competence is renamed to Naturality.

The last criterion which will be included into the scheme is the strategic competence. Skehan (1998) stated that this component concerns with metacognitive ability which automatically performed by each individual in setting the purpose of communication, planning a communicative act and implementing them. Furthermore, Mariano categorizes strategic competence into two categories; reduction and achievement strategies. Several skills included in reduction strategies are the ability in changing the topic and changing meaning. Achievement strategies include the ability in opening and closing a conversation, negotiate meaning, and maintaining communication and interaction.

Based on those categorizations, the form of the scheme can be seen on the table below.

	Area of Performance				
Scale	Grammatical Knowledge	Organization	Language Function	Naturality	Language Strategies
	Grammar Pronunciation Vocabulary Knowledge	Use of Cohesive Devices Sentence Organization	Ability to use language in accordance to its function	Appropriateness	Opening and Closing Meaning Negotiation

Table 11.4.1 deciding the scoring criteria

11.5 Deciding the Scale and Description of Criteria

The process of determining the scale and description in this research is formulated by forming a correlation between the objects observed, the aim of the program and class observation. The correlation will be based on the theory of Andrade (1997) and Underhill (1987).

The marking scheme scale is made by first writing the narrative descriptions for the highest and lowest scale. The description for the highest and lowest scale is determined by observing the highest and lowest students' performance. After observing the students' performance and correlating it with Underhill's statement on the amount of scale, it is decided that there will only be 3 scales included in the scheme. After the descriptions are set, the next step is deciding the amount of scale for the scheme. The end result of the scheme can be seen in the table 3 below.

	Area of Performance				
Scale	**Grammatical Knowledge**	**Organization**	**Language Function**	**Naturality**	**Language Strategies**
Exemplary	Students are capable of making complete sentences with generally accurate grammar. Mistakes are rare. Their pronunciation are generally clear and accurate. Mistakes are rare. Moreover, they can accurately use both old and new vocabulary.	Students can connect words and sentences using appropriate cohesive devices such as 'and','however', 'first', 'then', etc., and they are able to logically organize their utterance and rarely repeat them.	Students can use the language to express their ideas and feeling without assistance. Moreover, they can understand other people's utterance and respond to it accurately.	Students recognize the degree of formalities in a situation and use appropriate register in that situation. Moreover, they frequently use native stylistic devices in their utterance.	Students can open a conversation or utterance interestingly, effectively and flexibly. Students also knows what to do when they have to close a conversation or utterance. Moreover, they are not dominant nor passive in any interaction that they take part i
Satisfactory	Students are capable of making understandable sentences. The sentences are often incomplete and mistakes are frequent. The mistakes didn't interfere with the message. Their pronunciation are generally clear and accurate. Mistakes often occur to pronunciat	Students can connect words in simple sentences but have problems in connecting complex and compound sentences. They are also able to logically organize their utterance. However, they often tend to be repetitive.	Students often have difficulties finding the right expression to express their ideas and feeling without assistance. They can also understand others people's utterance. However, they often have difficulties in responding to it.	Students often misunderstand the situation in which they had the interaction. Students do not use proper style in their utterance. At times, they can repair their improper utterances.	Students do not vary their way in opening and closing a conversation or utterance. If they take part in an interaction, they are always too dominant. They will immediately translate difficult words to get the meaning across.
Novice	Students aren't even capable of making the simplest sentences. Mistakes are frequent and they interfere with the messages. Sentences are often organized only in groups of nouns. Mispronunciation occur frequently to both common and uncommon words. They oft	Students can even connect words in a simple sentence. Most of their utterance are illogical due to its bad organization.	Students always need assistance in expressing their ideas and feelings clearly and they cannot understand other people's utterance and respond to it accurately.	Students cannot recognize the degree of formalities in a situation resulting in frequent inappropriate use of register. Most of the times, they can only translate their mother tongue utterance into English.	Students cannot open or close a conversation or utterance. They always wait for the others to open or close a conversation. If they are involved in an interaction, they are always being passive.Moreover, they cannot get the meaning across. They do not wan

Table 11.5.1 Deciding the Scale and Description of Criteria

11.6 Testing the Scheme

The testing of the scheme was done using the steps proposed by Underhill (1987). The testing started with the explanation of the students' background and the program in which they join. It is then continued with the explanation on the reason for designing the test. After that, there are some explanations on the scheme and the detail of the scheme, such as the narrative description and area of performance. The next step is to test the scheme together which is immediately followed by a discussion on the testing. This process is called the initial testing. After several adjustments, the second phase of the test was initiated. This phase is called real-time testing. Finally, the results of the test will be tallied and the recapitulation will serve as the material for reconstruction.

11.7 Initial Testing

The main purpose of initial testing is to familiarize the teachers with the scheme along with the narrative description and criteria. The initial testing is done on the results of performance by Novi, one of the students in a TE Basic class. After being played twice, the assessment done by the 5 teachers can be seen on the table below. The criteria in the scheme were abbreviated in order to minimize confusion and make the recapitulation easier.

Teachers					
Criteria	Shinta	Ria	Dina	Yanna	Frantze
GK	Me	Me	Me	Me	Me
O	Ma	Me	Me	Ma	Me
LF	Ma	Me	Ma	Ma	Ma
N	Ma	Me	Me	Me	Me
LS	Ma	Ma	Ma	Ma	Ma

Legend:
GK   : Grammatical Knowledge
O   : Organization
LF   : Language Function
N   : Naturality
LS   : Language Strategies

Table 11.7.1 recapitulation of initial assessment

From the table it can be seen that the reliability factor has already existed in the criteria of Grammatical Knowledge and Language Strategies. This can be proven by the similarity in assessment for that criteria by the five teachers asked to assess the same area of performance. The assumption of reliability for the criteria is based on Brown's (2004) statement which stated that rater reliability can be measured by forming a correlation between raters; also known as inter-rater reliability.

11.8 Discussion on the Result of Assessment

Eventhough there are similarity, the dominant factor in the testing of the scheme is still in the difference of assessment for the remaining criteria. As seen in Table 4, there are still differences in the criteria of Organization, Language Function and Naturality. After a brief discussion, it is found that the difference is caused by the preference of some teachers doing the assessment. Moreover, the difference is also caused by the lack of understanding of the teachers in the criteria included in the scheme.

11.9 Real-Time Testing

The second phase of testing was done after the researcher is certain that the teachers are familiar with the criteria and the method of scoring. Different object is used to test the scheme; this is done to avoid the possibility of memorization in testing the performance. Recapitulation of the second phase of testing can be seen in the table below.

	Teachers				
Criteria	Shinta	Ria	Dina	Yanna	Frantze
GK	S	E	E	E	E
O	E	E	E	E	E
LF	E	E	E	E	E
N	E	E	E	E	E
LS	E	E	E	E	E

Table 11.9.1 recapitulation of real-time testing

From the second testing, it can be seen that the scheme has high reliability point. There is only one difference noted from the recapitulation of the scheme, which is in the criteria of Grammatical Knowledge. After investigation, it is found that the teacher misunderstood the inclusion of criteria and the description of performance. After being explained of the reason, the teacher agreed to change the assessment.

The results above show that the scheme contains qualitative reliability. This qualitative reliability still needs to be altered into quantitative reliability. Salkind (2004) suggested that the altercation can be made by weighing the amount of similar selection with the total amount of selections available. The illustration of this formula can be seen below.

Inter-rater reliability = the amount of similar selections/the total amount of selections

Based on the calculation above, the researcher has set the scale for the real-time testing reliability. The scale is to be used to determine the description of reliability of the designed scheme. Table 5 shows the scale of reliability and table 6 shows the interpretation of the scale on the value of reliability gathered from the real-time testing of the scheme.

Value	Reliability Description
0.8 – 1.0	Very strong reliability
0.6 – 0.8	Strong reliability
0.4 – 0.6	Weak reliability
0.2 – 0.4	Very weak reliability
0 – 0.2	Not reliable

Table 11.9.2 reliability scale

Criteria	Similar Selection	Total Selection	Percentage	Reliability Value
GK	4	5	80 %	0.8
O	5	5	100 %	1
LF	5	5	100 %	1
N	5	5	100 %	1
LS	5	5	100 %	1

Table 11.9.3 reliability scale interpretation

12. Conclusion

The designing of the scheme has been done using Bachman's concept of competence as the basis of the design, especially of the criteria. The analysis conducted on the scheme has shown that the scheme has a high value of reliability. This can be seen from the average value of the scheme which is 0.84

Moreover, even though the competence used is a familiar concept in ELT, some of teachers still have no knowledge on it. Therefore, the use of this scheme has to be initiated by familiarizing the teachers themselves to the concept of competence. This is to maximize the reliability of the assessment.

The results also show that the use of competence to assess the students' performance is acceptable as the performance is based on competence, and each part of competence as mentioned in the criteria of the scheme also exist in the students performance in doing a spoken task. Therefore, it can be said that the use of competence to assess students' performance has been verified.

References

Allen, Mary. 2003. *Using Scoring Rubrics. Students Learning Outcomes in the CSU*. Diakses 16 April 2007 from http://www.calstate.edu/AcadAff/SLOA/links/using_rubrics.shtml.

Andrade, Heidi G. 1997. Understanding Rubrics. originally published in *Educational Leadership,* 54(4). Diakses 23 November 2008 from http://www.middleweb.com/rubricsHG.html

Bachman, Lyle. 1990. *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bachman, Lyle & Palmer, Adrian S. 1996. *Language testing in practice: designing and developing useful language tests*. Oxford: Oxford University Press.

Brown, Douglas H. 1994. *Principles of Language Learning and Teaching (ed. ke-3)*. New Jersey: Prentice Hall Regents.

_____. 2004. *Language Assessment: Principles and Classroom Practices*. New York: Pearson Education.

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

Celce-Murcia, Marriane (peny.). 2001. *Teaching English as a Second or Foreign Language*. United States: Heinle & Heinle.

Cohen, Andrew D. 2001. *Second language assessment*. Teaching English as a second or foreign language. United States: Heinle & Heinle.

Dewaele, Jean-Marc. 2004. *The Acquisition of Sociolinguistic Competence in French as a Foreign Language: An Overview*. London: Birkbeck ePrints.

Figueras, Neus. January 2005. *Testing, testing, everywhere, and not a while to think*. ELT Journal Volume 59/1.

Heaton, J.B. 1975. *Writing English Language Test*. London: Longman.

Hedge, Tricia. 2000. Teaching and learning in the language classroom. Oxford: Oxford University Press.

Hesfiandani, Ninit. 2002. *An Assessment of the Effectiveness of English Teaching Materials: A Case Study at The Faculty of Engineering, Pakuan University, Bogor*. Tesis. Jakarta: Program Pascasarjana Universitas Katolik Indonesia Atma Jaya.

Hughes, Rebecca. 2002. *Teaching and Researching Speaking*. Great Britain: Pearson Education.

Hymes, Dell. 1972. *On communicative competence*. In Pride and Holmes 1972.

Joos, M. 1961. *The five clocks.* New York: Harcourt Brace Jovanovich.

Kaplan, Marshal A. and Knutson, Elizabeth. Spr. 1993. *Where Is The Text? Discourse Competence and Foreign Language Textbook*. Mid-Atlantic Journal of Foreign Language Pedagogy. Vol.1. 167 – 176.

Mariano, Luciani. June 1994. *Developing Strategic Competence: Towards Autonomy in Oral Interaction. Perspectives*, a Journal of TESOL-Italy - Volume XX, Number 1.

Mead, Nancy A. and Rubin, Donald L. 1985. *Assessing Listening and Speaking Skills*. ERIC Digest. Diakses 6 November 2008 from http://www.ericdigests.org/pre-923/speaking.htm

Mertler, Craig A. 2001. *Designing Scoring Rubrics for your Classroom. Practical Assessment. Research and Evaluation*, 7(25). Diakses 3 Maret 2004 from http://PAREonline.net/getvn.asp?v=7&n=25.

Moskal, Barbara M.. 2000. *Scoring Rubrics: What, When, and How. Practical Assesment, Research and Eevaluation*, 7(3), Diakses 21 Maret 2007 from http://PAREonline.net/getvn.asp?v=7&n=3.

_____. 2003. *Recommendation for Developing Classroom Performance Assessment and Scoring Rubrics. Practical Assesment, Research and Eevaluation*, 7(3), Diakses 16 April 2007 from http://PAREonline.net/getvn.asp?v=8&n=14.

Nunan, David. 1991. *Language Teaching Methodology: A Textbook for Teachers*. London: Prentice Hall.

Qin, Liu and Li, Wang. 2008. *A Portfolio Approach to Oral English Assessment for English Majors in China*. A paper presented in the 6th ASIA TEFL in Bali, Indonesia.

Salkind, Neil J. 2004. *Statistics for People Who (Think They) Hate Statistics*. Kansas: SAGE.

Skehan, Peter. 1998. *A Cognitive approach to language learning*. Oxford: Oxford University Press.

Stiggins, R. 2001. *Student-involved classroom assessment.* Upper Saddle River, NJ: Prentice-Hall, Inc.

Tomlinson, Brian. January 2005. *Testing, testing, everywhere, and not a while to think*. ELT Journal Volume 59/1.

Underhill, Nic. 1987. *Testing spoken language*. Cambridge: CUP.