AN EVALUATION OF THE SPEAKING ASSESSMENT TASKS USED BY AN ENGLISH TEACHER OF SMK NEGERI 4 MAKASSAR (A CASE STUDY)

Imran Trista Udin
Politeknik Informatika Nasional Makassar
Email: imrantristau@yahoo.com

Abstract

The objectives of this research were to find out (1) The description of the validity of speaking assessment tasks used at SMK Negeri 4 Makassar. (2). The description of the reliability of speaking assessment tasks used at SMK Negeri 4 Makassar. This research applied quantitative description. The population was the eleventh grade students of Tourism Department at SMK Negeri 4 Makassar, in 2011/2012 academic year. The population consisted of 30 students where the sample was taken by using purposive sampling technique. The research data were collected by expert validation, concurrent validity, content validity, face validity, test retest, reliability and inter rater reliability on speaking assessment task and analyzed by using SPSS 17.0 version to see the significant difference between test retest, concurrent validity and inter rater reliability. Then, the data on the expert validation collected by using expert validation and content validity and face validity collected by using native speaker analysis and comment. The results of the research were: (1) The reliability of speaking assessment tasks used at SMK Negeri 4 Makassar is high reliable and (2) the validity of speaking assessment tasks used at SMK Negeri 4 Makassar is valid. The researcher concluded that teaching material especially in speaking tasks is valid and reliable to teach at SMK (vocational school) where the English language teaching process is demanded to teach communicatively and functionally, so the SMK's students possess English competencies which are relevant to the job opportunities both in Indonesia and global setting.
Keywords: Evaluation, Assesment, Teacher

Abstrak

Penelitian ini bertujuan untuk melihat (1) gambaran validitas tugas berbicara siswa yang digunakan di SMK Negeri 4 Makassar (2) gambaran realibilitas tugas berbicara siswa yang digunakan di SMK Negeri 4 Makassar. Penelitian ini menggunakan metode kuantitatif. Populasi penelitian ini adalah kelas X jurusan Pariwisata SMK Negeri 4 Makassar, tahun akademik 2012/2013. Jumlah populasi penelitian ini sebanyak 30 siswa dimana sampelnya ditentukan dengan menggunakan teknik purposive sampling. Data materi belajar berbicara siswa dikumpulkan melalui validasi ahli, validasi perwajahan, dan validasi materi. Perbandingan validitas, tes ulang dan tingkatan realibilitas. Kemudian hasil kedua tes tersebut dianalisis dengan menggunakan program SPSS versi 17.0 untuk mengetahui perbedaan signifikansi antara guru dan peneliti. Sedangkan data validasi ahli, validasi perwajahan, dan validasi materi di ambil dari tanggapan penutur ahli. Adapun hasil penelitian ini menunjukkan bahwa (1) gambaran validitas tugas berbicara siswa yang digunakan di SMK Negeri 4 Makassar (2) gambaran realibilitas tugas berbicara siswa yang digunakan di SMK Negeri 4 Makassar. Peneliti menyimpulkan bahwa validitas dan realiblitas materi pembelajaran berbicara bahasa Inggris pada SMK (Sekolah Menengah Kejuruan) dimana proses pembelajarannya dituntut untuk didesain secara komunikatif dan fungsional, sehingga diharapkan

nantinya siswa-siswa SMK memiliki kompetensi bahasa Inggris yang sesuai dengan pangsa pasar kerja baik di Indonesia maupun dalam cakupan global.
Kata Kunci: Evaluasi, Penilaian, Guru

INTRODUCTION

Background Testing oral proficiency is become one of the most important issues in language testing since the role of speaking ability is more central in language teaching with the advent of communicative language teaching (Nakamura, 1993). As Bostwick and Gakuen (1995) state, assessment can be used to improve instruction and help students take control of their own learning. That is more likely to be accomplished when assessment is authentic and tied to the instructional goals of the program.

However, there are many difficulties involved in the construction and administration of any speaking assessment. There is a great discrepancy between the predominance of the Communicative Approach and the accurate measurement of communication ability (Hughes, 1989). With the widespread adoption of communicative language teaching (CLT) in ESL countries (Hartley and Sporing, 1999), CLT is replacing the traditional grammar-centre, text-centre, and teacher-centre methods in Vocational High School. With training in CLT offered by in-service teacher education programs, Vocational High School English teachers have been trying to expand CLT in their classrooms (Li, 1998).

Although many English teachers at vocational high school are interested in CLT, communicative assessment has received little attention. If it is important to know if a person can speak a second language, then it should be important to test that person's speaking ability directly (Jones, 1977). Despite the interdependence of communicative teaching and communicative assessment (Bachman, 1990), speaking assessment in the Vocational High School does not assess students' oral proficiency from the perspective of language use and communication (Li 1998).

As Nagata (1995) pointed out, rote memorization of text dialogs has been a common practice for speaking assessment at vocational high school. It seems that vocational high school English teachers do not concern themselves much with matters of validity and reliability in relation to speaking assessment (Li, 1998). However, the need for classroom teachers to be equipped with some measurement tools to evaluate student's oral proficiency is becoming more and more important.

Language assessment in the vocational high school oral proficiency is becoming more and more important (Nagata, 1995). Speaking assessment has become a vital part of all the examinations in vocational high school in Indonesia; every school is required, by the Ministry of Education and Culture, to perform students' speaking assessment at least once each year. The schools referred to in this study take a speaking assessment twice a year. Vocational high school English teachers, however, find it difficult to assess students' oral proficiency in a way which reflects authentic interaction (Li, 1998). Thus, there is a contradiction between their communicative language teaching and their assessments of speaking which do not reflect authentic oral interaction.

In this research, the writer wants to discuss an evaluation of English teacher-speaking assessment tasks, and based the facts stated in previous explanation in the background, the writer formulates some research question in order to find out the reliability of speaking assessment tasks used at SMK Negeri 4 Makassar and to describe the validity of speaking assessment tasks used at SMK Negeri 4 Makassar?

The finding of this research will contribute to the larger body of knowledge for prepare a test for the students and additional information to teaching research especially those dealing with the validity and reliability test. Besides that, the writer hopes that this finding can help the teachers to organize, develop, or select the good tests for their students. Then, this finding also can improve the development of a good quality of test.

This research is limited to the evaluation of the teacher speaking assessment tasks used at SMK Negeri 4 Makassar. The speaking tests evaluated are the tests which are made by the English teacher at SMK Negeri 4 Makassar. The tests analyzed are the tests which are given to the students at the school who enrolled in the 2011/2012 academic year. The observation will focus on speaking materials that cover; Introductions self and Greeting and leaving taking.

Related Literature

This part deals with the previous related studies, theories or concept related to the thesis' topic, resume, and conceptual framework. The principal areas to be discussed are communication and communicative competence, communicative Language Teaching (CLT), speaking assessment, and the teachers' perceptions of speaking assessment task.

There are some findings that are related with this research. They are summarized as in the following that serve as an overview to this current research. Alauddin (2002) reported in his research finding on analysis of the teacher-made multiple choice English test for the students of three school in Polewali that the teacher-made multiple choice English test had many errors in the item construction and the item materials, it's caused by the lack of preparation. He also reported that many poor item were found in the test. Moreover, the test items were not revised and tried out before using. It indicated that the teacher-made multiple choice English test construction for the students of SMAs in Polewali sub-district did not follow the standard criteria of construction. Whereas the result in item analysis of test showed that the test item was difficult for students, and also the average of the discrimination index was not satisfactory. The reliability of the test is normal in the level of teacher-made test. At last based on the procedures of the analysis, he found that the test items for the second year students were not valid for the future use. Pammusu (2004) analyzed on the teacher-made grammar tests of State SLTPs in Je'neponto. He reported that the teacher-made grammar test tested at SLTP in Je'neponto can be used, because they have adequate discrimination power, validity, reliability, and distracter function, even though, the average of difficulty level of the test construction and need to be revised, the test can be used. Another important research finding on the analysis of the teacher-made test thought was done by Jusni (2009) she analyzed the English teacher item used at SMA Negeri 3 Makassar. And reported that the English test items used in grade XI of SMA Negeri 3 Makassar, were infeasible. This fact simply means more than a half of the English test items used in grade XI of SMA Negeri 3 Makassar was unable to deal with the student's ability. Whereas, the English test items used in grade XI of SMA Negeri 3 Makassar were invalid and were unable to deal with the standard of validity index required by trustworthy test item; the ability of this item used in grade XI of SMA Negeri 3 Makassar were reliable. She suggested that in creating the best item, the teacher should give more attention to the items analysis which is the feasibility, validity and reliability. Then, the test item should be reviewed and tried out before being used.

From the previous of related finding there are some reason why the test is negative for the students, such as, lack of preparation, less attention to the items, less of review, and not try out of the test.

Related to the previous findings above, the writer need to conduct a research an evaluation of teacher-speaking assessment tasks focus on mid test to know how effective the materials which have been taught and the students performance in mastery the materials, and also critical analysis on each item of test thus the item is acceptable or unacceptable.

Method

The method used in this research will be descriptive quantitative. Descriptive method will be used to determine and describe the way things are, test analysis. Descriptive qualitative method is the data analysis using statistical calculation. The result of the descriptive statistics method is the empirical judgment on the quality of the test based on the statistical calculation; in this case the writer used computerized calculation of the data obtained. The result of the research was intended to know the quality of the teacher-speaking assessment task at SMK Negeri 4 Makassar.

The Research Design

The research will use descriptive quantitative design. This is to describe validity and reliability of speaking assessment task.

The Research Variables

This research has only one focus that becomes the object of the research. The variable of this research is the English teacher- speaking test; objective tests. This variable consists of sub-variables such as quantitative description of the test. They are (1) validity, and (2) reliability of the test.

The Operational Definition of the Variable

Speaking assessment task is any written document given by teacher. Validity is defined as the ability of the test items to measure what is supposed to measure, and Reliability is defined as the ability of the test items to be consistent in terms of measurement.

Population of the Research

The population of this research is the English teacher speaking assessment tasks used for the students at the SMK Negeri 4 Makassar in the 2011/2012 academic year. The English teacher-speaking assessment tasks are used for the students in grade II.

Sample of the research

In this research, the researcher use purposive sampling technique where only one class will take as the sample of this research, namely class 1 TNT 1. This because the students from class 1 TNT 1 will get on the job training in industries for one semester or six months. So, the total number of samples will 24 (twenty-four) students.

Procedure of Collecting Data

In collecting data the writer carry out the following procedures; 1) Collects the speaking material test to be given to the students. 2) Observes the speaking assessment. 3)Collects the data of the students' answers; and 4) Analyzed the tests and the students' answer.

The Data Analysis Method

The analysis used in analyzing the data was quantitative analysis. The quantitative analysis was intended to get information about the item, whether the items have been effective or not. Item result analysis can be used to revise the poor

items or to choose the items based on certain characteristics to be taken as qualified test items. The quantitative analysis result of the test and the item characteristics are follows

Test Validity

The validity of a test is the extent to which it measures what is supposed to measure and nothing else. There are several kind of validity, face validity, content validity are to be the internal validity, while concurrent validity and expert validation are to be external validity.

Face validity

- If a test item looks right to other testers, teachers, validators, and testee, it can be described as having face validity. It is, therefore, often useful to show a test to colleagues and friends, face validity is often determined impressionistically

Content validity

The test should be so constructed as to contain representative sample of the course, the relationship between the speaking task and the course objectives always being apparent, when embarking on the construction of a test, the teacher should draw up a table of specification, describing in a very clear precise terms the particular language skills areas to be included in the test, some specialists make no distinction between content and face validity and consider them to be synonyms

Concurrent validity

To validate this kind of empirical, criterion-related validity, one administers a recognized, reputable test of the same ability to the same persons concurrently or within a few days or months of the administration of the test to be validated. Score of the two different tests are then correlated using some formula for the correlation coefficient, and the resultant correlation coefficient is reported as a concurrent validity coefficient.

Expert validation

Expert validation intended to describe the expert analysis to the speaking material.

Reliability

Reliability is a measure of accuracy, consistency, dependability, or fairness of score. There is some common method for computation of reliability, which involve statistic calculation, correlation and variances of score.

Test-retest method

This is calculated by means of product moment correlation of two sets of scores for the same person, the same test is administering to the same students following an internal of no more than two weeks.

FINDINGS

This research consists of two sections. They are findings and discussions of the research. The results are presented in the form of data description and the result of data analysis covered (1) Validity and (2) Reliability.

The findings are reported in this section are based on the procedure of data collection in the previous chapter. The findings of the research deal with data analysis which covered (1) Validity and, (2) Reliability.

Validity

To see whether the speaking assessment task given by the students valid or invalid, the researcher used a validity expert and concurrent validity as external validity and as internal validity they are the face validity and content validity.

External validity

Expert validation

The expert validation consists of two validators, and all validator assumed that the test is valid but need bit revisions.

Concurrent Validity

Concurrent validity takes a score from the first and second semester, to see the reputable test of the same ability to the same students concurrently of the or within a few months of the administration of the test to be validated. And the result of the concurrent validity first and second semester by used formula (pearson product moment) in SPSS 17 got 0,68 it mean the correlation is strong.

Internal Validity

Face validity

In internal validity the researcher took an impression from native speaker from America, the native speaker give his impressions about the presentation of the students in front of classroom when they are presented the speaking assessment tasks, and the researcher found the impression from the native speaker that the students are speaks with clear pronunciation, most of them use simple sentences (subject+ verb+ object) and most of the students seldom use a connectors like next, after that and etc.

Content validity

In content validity the researcher still took a native speaker, to analyzed the contents sample of the speaking assessment task, the relation between the test teaching material, and the result of the content validity shows that there are some sentences need to be revised before explain to the students, like what's news? Should, be what's new? And the response is nothing much not just fine thanks.

Reliability

For measure the Reliability of the speaking assessment task used by teacher, the researcher applied test – retest approach. This approach applied by gave a speaking assessment task twice to the students, and the scores calculated by use formula (pearson product moment) in SPSS 17 and the researcher found that the reliability of the students speaking assessment task with test – retest approach found 0.78 it means the reliability is strong.

DISCUSSION

This part described the result of the analysis tests which deals with the interpretation of the findings taken from the previous quantitative analysis.

Expert validation

Based on the data analysis of the expert validation, the researcher finds that the validator has different perception of speaking materials. The researcher described them one by one.

Based on the analysis of validator A, he assumed that aspect number one is the test, from the four statements, statement a – c the validator A gave score 4 (very good) while in statement d the validator A gave point three (good). In number two that is construction aspetc, in construction aspect there are three statements.

Statements a and c the validator A gave score four (very good). While in statement b validator A gave score three (good).

In aspects number three that is language, in language aspect there are three statements, for statement b and c validator A gave score four (very good), while in statement A as the validator A give score three (very good), and the last number of aspect is time, in time aspect there are two statements, statement a – b the validator A gave score 4 (very good). And the last of validation the Validator A give a comment that please avoid use an unusual sentence like how do you do, from the analysis above validator A gave score three (good) or the test is valid but need little revision.

In analysis of validator B, he assumed that aspect number one is the test, from the four statements, statement a – d the validator B gave score 4 (very good). In number two that is construction aspect, in construction aspect there are three statements. Statements b and c the validator B gave score three (good). While in statement a validator B gave score four (very good).

In aspects number three that is language, in language aspect there are three statements, for statement a - c validator B gave score three (good). And the last of validation the Validator B gave a comment that some task is wrong. For the part of greeting, "where are you going?" just suit for Japanese people. From the analysis above validator A gave score three (good) or the test is valid but need little revision.

Concurrent Validity

To validate this kind of empirical, criterion-related validity, one administers a recognized, reputable test of the same ability to the same persons concurrently or within a few days or months of the administration of the test to be validated. Score of the two different tests are then correlated using some formula for the correlation coefficient, and the resultant correlation coefficient is reported as a concurrent validity coefficient. In concurrent validity, the researcher took the result of semester task, first semester and second semester, and after calculated by used SPSS 17 the result show that the correlation between first semester and second semester is 0.68 it means the correlation is strong.

Correlations

		S1	S2
S1	Pearson Correlation	1	.680**
	Sig. (2-tailed)		.000
	N	30	30
S2	Pearson Correlation	.680**	1
	Sig. (2-tailed)	.000	
	N	30	30

**. Correlation is significant at the 0.01 level (2-tailed).

Face validity

If a test item looks right to other testers, teachers, validators, and testees, it can be described as having face validity. It is, therefore, often useful to show a test to colleagues and friends, face validity is often determined impressionistically. In face validity the researcher took an impression from the native speaker, before the native speaker gave the impression about the student presentation about speaking assessment

task, the researcher showed a video, and the content of the video is the student presentation about speaking assessments task, and after saw the video the native speaker gave comment about the student presentation in speaking assessment task, she said that the students speak clearly, they just used a simple sentence they are Subject +Verbs + Object, and another comment she said the students seldom use a connectors like next, after that, therefore and etc.

Content validity

The test should be so constructed as to contain representative sample of the course, the relationship between the speaking task and the course objectives always being apparent, when embarking on the construction of a test, the teacher should draw up a table of specification, describing in a very clear precise terms the particular language skills areas to be included in the test, some specialists make no distinction between content and face validity and consider them to be synonyms, while the native speaker saw the video, the researcher also showed the teaching material relate to speaking assessment task, and the native speaker commented some material that need to be revised based on the own language, like in teaching material the teacher write " what's news?" and the response just fie thanks, and the analysis of the native speaker it should be "what's new?" and the response is "nothing much, what's new with you?" and another analysis the teacher write in teaching material "I'll talk you later" and the response is "sure, see you later" and the analysis of the native speaker is the response should be "sure talk you later"

Reliability

In applied reliability in speaking assessment task, the researcher use test-retest, this is calculated by means of product moment correlation of two sets of scores for the same person, the same test is administering to the same students following an internal of no more than two weeks. After the researcher calculated the result of the first and the second time of the test, the researcher found 0.96 it means the correlation between the first test and the second test is very strong, the data analyzed by use formula in SPSS 17.

Correlations

		X1	X2
X1	Pearson Correlation	1	.967**
	Sig. (2-tailed)		.000
	N	30	30
X2	Pearson Correlation	.967**	1
	Sig. (2-tailed)	.000	
	N	30	30

**. Correlation is significant at the 0.01 level (2-tailed).

CONCLUSIONS

Based on the research findings and discussion in the previous chapter, the researcher comes to the following conclusions referring to the research question and the objectives of the research.

Every validator assumed and gave the score three (good or the test is valid but need bit revision) to the speaking assessment task used by an English teacher in SMK Negeri 4 Makassar. The finding of tabulated data shows the correlation of concurrent validity, which cover the first and the second semester shows that 0.68 it means the correlation is strong.

The face validity describes as having face validity, and the analysis of the native speaker to the face validity after saw the video, said that most students speak clearly.

Content validity is representative sample of the course, the relation between the speaking assessment task and the course, to see whether the course and speaking relate each other, the researcher asked the analysis of the native speaker to the correlation, and she assumed that the material still need bit revision. The finding of tabulated data showed the reliability of the speaking assessment task, the researcher applied test-retest method to see the correlation between the first and the second test, the result is 0.96 it means the correlation is very strong.

BIBLIOGRAPHY

Alauddin, (2002). The analysis of the Teacher-made test multiple choice English test for the students of state SMASs in Polewali Sub-District, Unpublished thesis" Pascasarjana UNM Makassar

Arikunto, Suharsimi. (2003). Dasar-dasar pemikiran pendidikan. Edisi Revisi: Jakarta: Bumi Aksara.

Bachman, L. and Palmer, A. S. (1984). Some comments on the terminology of language testing. In: Rivera, C. (ed.), Communicative Competence Approaches to Language Proficiency Assessment: Research and Application. Clevedon: Multilingual Matters, 34-43.

Bailey, Kathleen M. and Savage, L. (1994). New Ways in Teaching Speaking. Cambridge: Cambridge University Press.

Backhoff, Eduardo. (2000). The level of difficulty and discrimination power of the basic knowledge and skill examination (EXHCOBA). Vol. 2. Online http://radie.uabc.mx/vol2no1/contents-backhoff.html. Retrieved: January 25, 2013 at 21.00

Clark, Herbert H and Clark, Eve V. (1977). Psychology and Language: An Introduction to Psycholinguistics. USA: Harcourt Brace Javanovich, Inc.

Cronbach, L.J. (1971). Test Validation. Dalam R.L Thorndike (Ed) Educational Measurement (2nd ed). Washington, DC: American Council on Education.

Ebel, Robert L. (1965). Essential of Education Measurement. The second Edition. New Jersey: Prentice Hall, Inc, Englewood Cliffs.

ETS. (2006). Examinee Handbook. Jakarta: PT. International Test Center.

Fulcher, Glenn. (2003). Testing Second Language Speaking. Hong Kong: Pearson Education Limited.

Gay, R.L. (1987). Educational Research: Competencies for Analysis and Application (Second Edition). New York: Merrill Published Company.

Gronlund, Norman E. (1971). Measurement and Evaluation in Teaching. New York: Macmillan Publishing Company.

Hughes, A. (1989). Testing for Language Teachers. Cambridge: Cambridge University Press.

Harmer, Jeremy. (1991). The Practice of English Language Teaching. New York: Longman.

Heaton, J.B. (1989). Writing English Language Test. USA: Longman Inc.

Jones, R. L. (1977). "Testing: A Vital Connection". In Philips, J. K. (ed), The Language Connection, 237- 65. Skokie, 3: National Textbook Co.

Jabu, Baso. (2008). English Language Testing. Makassar: the UNM Published

Kizlik, Bob. (2009). Measurement, Assessment, and Evaluation in Education. Online http:www.adprima.com/measurement.htm. Retrieved: January 25, 2013 at 21.00

Li, D. (1998). "It's always More Difficult Than You Plan and Imagine": Teachers' Perceived Difficulties in Introducing the Communicative Approach in South Korea. TESOL Quarterly, 32, 4.

Loannove, Sophie., Geourgio, and Pavlov. (2003). Assessing Young Learners.Oxford: Oxford university press.

Nakamura, Y. (1993). Measurement of Japanese college students' English Speaking ability in a classroom setting. Unpublished doctoral dissertation, International Christian University, Tokyo.

Parrott, Martin. (2002). Tasks for Language Teachers. UK: Cambridge University Press.

Pammusu. (2004). An analysis of teacher-made test grammar test of state SLTP in Je'neponto Regency Unpublished thesis" Pascasarjana UNM Makassar

Richards, Jack; John Platt, and Heidi Weber. (1985). Longman Dictionary of Applied Linguistics. London: Longman Group UK Ltd.

Saenong, (2011). Critical analysis of English teacher-made test used at three state SMAs in Makassar, Unpublished thesis" Pascasarjana UNM Makassar

Surapranata, Sumarna. (2005). Panduan Penulisan Tes Tertulis. Bandung: Remaja Rosda Karya.