

Lecturers' Considerations in Developing Language Test

(A Study of English Education Lecturers at English Education Departments in Parepare)

Nur Fadillah Nurchalis

Ibrahim Ieman

Amzah Selle

Institut Agama Islam Negeri (IAIN) Parepare

Address: Amal Bakti No.8, Soreang Kota Parepare, Sulawesi Selatan 91131

nurfadillahnurchalis@yahoo.com

Abstract: This research aims to find out (1) the neglected variables of characteristics of a good test which is done by lecturers in developing their test, and (2) the ways the lecturers develop their own test in relation to the neglected variables. This research is explanatory mix method design. There are two variables in this research. They are lecturers' consideration and test development. The population of this research is 33 lecturers in which 18 from IAIN Parepare and 15 from UMPAR. The samples are taken by random technique in which 30 samples. The instruments of this research are questionnaire, interview and documentation. This research uses percentage technique to analyze the data. The results of this research show that (1) here are the sequences of the neglected variables in developing language test: reliability, appropriateness in difficulty, clarity, comprehensiveness, validity, transparency, appropriateness in time, and economic, and (2) From three interview sections, this research finds differences of the way lecturers develop their test. They develop their test through challenging, creative and spontaneous test.

Keywords: *Language testing, consideration, test development*

1. Introduction

Educational process, formally, at schools is more well-known as teaching and learning process. Teachers are always associated with teaching activity. However it is only the one of series of the real educational process. Government has pointed out teachers and lecturers' roles in education in a regulation number 14 in 2005.

Language teachers are often encouraged to use creative teaching in the classroom (Aswad 2017). Teachers do not only teach students at school, but also educate, guide, train, assess, and evaluate them, while lecturers are professional educators who transform, develop, spread knowledge, technology, and art through education, research, and services to society. The regulation above shows that there are some responsibilities that both teachers and lecturers undertake in education. However, term "through education" means that teachers' responsibilities above also involve in lecturers' responsibilities. Those all are integral part or a unity of educational system which lead to the growth of competence. Teaching at schools or universities is not only a routine that is systemic, but also it prefers to achieve certain objectives. To know whether it is achieved or not, teachers or lecturers should conduct an evaluation. Evaluation gives information to what extent a program works or how the progress is. Related to educational evaluation, a regulation Number 20 Year 2003 explains about the purposes of evaluation.

Evaluation aims to control educational quality nationally to all who and which is associated with it such as students, institution, and educational programs. Another regulation above points out that

evaluation is a vital part in education that both teachers and lecturers should realize. Evaluation is not the last step in education, but it becomes a starting point to the future program. Educational quality is reflected by its evaluation. Cross (1973: 5) stated that Evaluation was a process which determined the extent to which objectives had been achieved.

Evaluation should be done systematically to find out a comprehensive information about students' development in learning process. Evaluation can be done through test or non test such as observation and interview. Nevertheless to check students' cognitive, a test is a common instrument for teachers and lecturers.

To improve students' quality, lecturers should give academic best services to students. So that as the output of learning, students who will be the alumni will be qualified. They play their roles as well in their society. One of indicators to see students' success in learning at university level is GPA. To find out the GPA, lecturers conduct an evaluation or measurement. While to find out a good result, the instrument of measurement should be good too.

In the end of teaching, lecturers are recommended to construct a test. It is well known as teacher-made test. Constructing test is not easy for some lecturers. However, that is a challenge for them to construct and develop a test. This is one of ability that lecturers should learn because it influences learning quality. Testing is also sometimes denied. Lecturers are often so busy with their other activities such as preparing media for teaching and conducting some activities to raise their "weight score"

for certification.

Lecturers in university level and teachers in both elementary and high school level have different challenges in constructing a test. In both elementary and high school level, teachers have a community, *Kelompok kerja Guru (KKG)* and *Musyawarah Guru Mata Pelajaran (MGMP)*. There they discuss and share together about many things such as problems that they face in the classrooms, suitable methods to use, and test that they gave to students. In the contrary, lectures at university level should construct the test by themselves. It is certainly more challenging and the authority level of the test is higher.

When the researcher was in college years, she conducted a free observation. She found some problems in teacher-made tests. Some lecturers did not revise their test every final semester. So the test takers had already known the test content from their senior. It made information about students' progress, result of evaluation, bias. Besides, the test also commonly was monotone. It was always in multiple choice items or essay while there were many kinds of items that could be varied in constructing test.

In addition, a research which was done by Widowati (2011) found that the teacher-made try-out test of UAN 2010/2011 of junior high schools in Malang had good face validity and content validity. It was also good in terms of the reliability, item validity, discrimination index, and item difficulty, but the distracters of the items were not good. The teacher-made test above generally reflected that the way the teacher constructed or developed this test was good.

Not all teacher-made tests are good; however they may come up in variety. Every

lecturer has their own considerations in constructing or developing a test. Therefore in this research, the researcher would like to know lecturers' considerations in developing a test and how they develop their hand-made test.

Based on the background above, the researcher formulates problem statements as follow: 1) What are the neglected variables of the characteristics of a good test by lecturers in developing their test? , and 2) How do the lecturers develop their own test in relation to the neglected variables.

This research is designed for explanatory mixed method design in which quantitative method is followed by qualitative method. To find out the first problem statement above, the researcher uses a quantitative method while to find out the second problem statement, the researcher uses a qualitative method.

There are two variables in this research. They are lecturers' considerations and test development. Lecturers' considerations here mean considerations based on the characteristics of a good test which are combined from theories of P.Harris, Junaedi and Ramadhan. They are validity, reliability, comprehensiveness, and appropriateness in difficulty, clarity, appropriateness for time, transparency, and economy. Test development here is that the ways of the lecturers develop their test in relation to the neglected variables of characteristics of a good test, and why they neglect them while prefer others. Test here is a test which is made by English lecturers at English Education Departments in Parepare when final examination. Lecturers here are English lecturers at

English Education Departments at both UMPAR and IAIN Parepare.

Population in this research is English lecturers at English Education Department at two universities in Parepare, namely IAIN Parepare and UMPAR. There are 18 lecturers at IAIN Parepare and There are 15 lecturers at UMPAR, so that the number of population is 33 lecturers. To fulfill a quantitative analysis, a research should use probability sampling. Therefore in this research the researcher takes 30 lecturers as samples by using random technique sample. For interviewed respondents, there are three lecturers covering both universities.

The instruments of this research are questionnaire, non-structural interview and documentation. Questionnaire aims to answer the first problem statement related to the lecturers' considerations in constructing their test. While non-structural interview and documentation aim to answer the second problem statement related to the way they develop their test.

The researcher distributes questionnaire to lecturers to find out the neglected considerations, characteristics of a good test, when they construct their own test. The researcher analyzes the questionnaire which contains lecturers' considerations namely validity, reliability, comprehensiveness, appropriateness in difficulty, clarity, appropriateness for time, transparency and economy by using percentage technique. Through percentage technique, the researcher finds the neglected characteristics of a good test which is done by lecturers in constructing their test. The researcher interviews the lecturers related to the reasons of their considerations and the ways they develop

their own test.

The data from questionnaire are analyzed on the degree of lecturers' considerations by using Rating Scale. For qualitative data, they are analyzed by using "Flow model" proposed by Miles and Huberman (1984). This model has four components. They are 1) data collection, 2) data reduction, 3) data presentation, and 4) drawing conclusion or verification.

2. Findings And Discussion

This research concerns on three problem statements. The findings of them will be discussed in the following part orderly.

There are eight variables that lecturers should consider in developing their own test. Those are: (1) Validity, (2) Reliability, (3) Comprehensiveness, (4) Appropriateness in difficulty, (5) Clarity, (6) Appropriateness in time, (7) Transparency, and (8) Economy. To know what the neglected are, the researcher will present the detail percentage of each variable. In the end of this part the researcher orders those variables from the most neglected variable to the considered variable.

A. Validity

In this section, the researcher found data regarding which part of validity of the test that is mostly neglected by the English education lecturers in Parepare and how many percent of the lecturers that neglect the variable. Besides, explanation in detail related to each item that represents valid variable will be shown. It is described in the following table.

Table 1. Frequency and Percentage of Validity Variable

Item	1		2		3		4		5		Total	
	Test Development based on the memory		Oral Test for non Speaking Subject		Attitude in Oral Performance		Neglecting test Specification		Empirically Valid Analysis			
	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)
always	0	(0)	16.7	(5)	6.67	(2)	6.67	(2)	10	(3)	8	(12)
often	3.33	(1)	0	(0)	23.3	(7)	13.3	(4)	16.7	(5)	11.3	(17)
Sometimes	16.7	(5)	30	(9)	43.3	(13)	13.3	(4)	23.3	(7)	25.3	(38)
seldom	30	(9)	23.3	(7)	23.3	(7)	36.7	(11)	28	(8)	28	(42)
Never	30	(15)	30	(9)	3.33	(1)	30	(9)	27.3	(7)	27.3	(41)
Total	100	(30)	100	(30)	100	(30)	100	(30)	100	(30)	100	(30)

Table 1 above shows that in there are 16.7% (5) of lecturers that always used oral test to other subjects except speaking subject. There are 23.3% (7) of lecturers that often concerned to students' attitude and other aspects in testing students' oral performance. There are 43.3% (13) of lecturers that also sometimes concerned to students' attitude and other aspects in testing students' oral performance and there are 36.7% (11) of lecturers that seldom neglected making their test specification. It means that they often made their test specification while there are 50% (15) of lecturers that never arranged their test based on the material that they remembered without considering indicators of learning. It means that they considered to arrange the test orderly based on the syllabus. Above all the items, the total percentage of consideration is higher for "Seldom" (28% (42) lecturers) and "Never" 27.3% (41) lecturers) level. It indicates that the lecturers consider a valid test enough.

These data are supported by an

interview section between the researcher and one of lecturers (Responden 1). Here is the conversation:

Researcher said: "Sir, what is your consideration to use oral test for Reading subject?"

Responden 1 said: " Actually, it depends on what class that I face. For freshman and sophomore level, I use oral test for them. It aims to train them speak more. Different from it, for junior and senior level, I use written test because they have been good at speaking. I prefer using kinds of test in which I found students' weakness. It purposes to challenge them to improve and motivate them to master their weakness".

Based on the statement above, the researcher finds that sometimes lecturers use unusual or different test methods as other considered variables. However, those unusual test methods that they consider are not meaningless. They consider them regarding their advantages for the students. In addition, information from Respondent 2 strengthens the finding above. Here is

his statement:

“Yes, There are many factors that I consider in Speaking test. I asked the students to make a video in group in which they make a conversation among them related themes that I had provided. Automatically, their creativity in making video, their expression in speaking, etc are being my consideration”

In a Speaking test, in Heaton theories, aspects to measure students’ speaking ability have five criteria, namely fluency, accuracy, vocabulary, content, and pronunciation. Nevertheless, the fact shows that other aspects also influence lecturers in constructing their test. It indicates that the test does not measure well what it should be measure. However, He added

Table 2. Frequency and Percentage of Reliability Variable

Item	1		2		3		4		5		Total	
	Preference to fewer question		Preference to face validity		The use of test		Preference to subjective test		Empirically reliable test			
	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)
always	30	(9)	13.3	(4)	23.3	(7)	33.3	(10)	10	(3)	8	(12)
often	46.7	(14)	33.3	(10)	30	(9)	10	(3)	16.7	(5)	11.3	(17)
Sometimes	13.3	(4)	30	(9)	30	(9)	10	(3)	23.3	(7)	25.3	(38)
seldom	10	(3)	13.3	(4)	10	(3)	26.7	(8)	28	(8)	28	(42)
Never	0	(0)	10	(3)	6.67	(2)	20	(6)	27.3	(7)	27.3	(41)
Total	100	(30)	100	(30)	100	(30)	100	(30)	100	(30)	100	(30)

Table 2 above shows that in there are 33.3% (10) of lecturers that always made subjective test than objective test. There are 46.7% (14) of lecturers that often preferred making fewer questions than more questions in a test. There are 30% (9) of lecturers that sometimes considered if their test met face validity, they did not need to do a reliability test and used the test only once to the same students. While there are

that the validity of communicative test was dependent on the test-creator’s understanding and definition of the term. If a validity of a test relies on the test-creator, it means that all of the communicative tests are probably guaranteed their validity. Nevertheless the lecturers or the teachers should not rely on this statement. They should make sure that their tests deal with minimally one kind of validity of the test such as face validity.

B. Reliability

In this section, the researcher finds which part of reliability of the test that is mostly neglected by the English education lecturers in Parepare and how many percentage lecturers consider it. It is described in the following table.

26.7% (8) of lecturers that seldom preferred to use subjective than objective test and

neglected doing a reliability test empirically. Besides, there are 26.7% (8) of lecturers that never neglected an analysis of reliability empirically. It means that they often did empiric analysis of reliability. If the number of percentage in “Always” and “Often” are combined, there is higher number lecturers who do not consider the

reliability of the test as well than who do it.

The findings above are supported by information from interview sections between researcher and respondents.

Researcher said: *“What is your reason why do you prefer subjective test to objective test?”*

Respondent 1 said: *“Objective test is too easy to test students’ language ability because students have potency to guess the answers. Besides, it is easy for them to cheat each other”*

Here is other statement from respondent 2. He said *“Speaking is skill subject. If the students only answer A-B-C, that is not measuring skill, but it prefer to knowledge. Therefore a speaking test should reflect skill”*

Those statements above show that for university level, objective test is less challenging for students. However, a challenging test should not neglect the quality of the test. Heaton (1988) said that objective test overcomes scoring the test of marker reliability, but subjective test is sometimes faced with it. Objective test is also challenging. It depends on level of familiar word of the test, distracters of the test, and etc. A challenging test should not deny reliability of the test.

Table 3. Frequency and Percentage of Comprehensives Variable

Item	1		2		3		4		5		Total	
	Preference to fewer question		Preference to face validity		The use of test		Preference to subjective test		Empirically reliable test			
	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)
Always	16.7	(5)	43.3	(13)	40	(12)	0	(0)	23.3	(7)	24.7	(37)
Often	3.33	(1)	26.7	(8)	23.3	(7)	0	(0)	3.33	(1)	11.3	(17)
Sometimes	0	(0)	10	(3)	13.3	(4)	13.3	(4)	33.3	(10)	14	(21)
Seldom	0	(0)	3.33	(1)	16.7	(5)	33.3	(10)	20	(6)	14.7	(22)
Never	80	(24)	16.7	(5)	6.67	(2)	53.3	(16)	20	(6)	35.3	(53)

In addition, some lecturers said that they do not need to do a reliability test if their test deals with face validity. Here is respondent 3 statements:

“It only takes time. The most important is that it is relevant to learning indicators in syllabus. Generally, only few lecturers who do it because lecturers undertake many things to do. Moreover, for those who have been certified. They should do their research, other activities to increase credit point and etc”

The statement above indicates that lecturers seem have no time to examine their test reliability. Too many jobs or activities that is more important than it. Nevertheless in the questionnaire some of them admit that they do it empirically. All in all, based on the total percentage, this variable is the most neglected factor among others. It is neglected 15.9% to consider.

C. Comprehensiveness

In this section, the researcher finds which part of comprehensive variable of the test that is mostly neglected by the English education lecturers in Parepare and how many percents of lecturers consider each item. It is described in the following table.

Total	100	(30)	100	(30)	100	(30)	100	(30)	100	(30)	100	(150)
--------------	------------	-------------	------------	-------------	------------	-------------	------------	-------------	------------	-------------	------------	--------------

Table 3 above shows that there are 43.3% (13) of lecturers that always took only some materials as representative of the test. There are 26.7% (8) of lecturers that often also took some materials as representative of the test. There are 13.3% (4) of lecturers who sometimes arranged the order of the materials randomly and considered to take many materials into their test even though they had fewer meetings with their students. Further, there are also 33.3% (10) of lecturers that seldom considered to take many materials into their test even though they had fewer meetings with their students, while there are 80% (24) of lecturers that never put any kinds of materials that they did not teach to the students. It means that they took test materials based on their teaching materials. If both “Always” and “Often” percentage above are added namely 36%, it is still lower than “Seldom” and “Never” percentage namely 50%. It indicates that

the comparison between lecturers who consider and not is not too many.

Some respondents’ statements support those findings above. Here is one of them:

Respondent 1 said:

“Well, I dense my teaching material into a test. Therefore there are only few numbers in my test. Thing that I emphasize is that the test should contain materials that we (lecturers and students) have discussed”

The statement above shows lecturers’ preferences and emphasis to what they teach to the students. This variable is neglected to consider 13.4 %.

D. Appropriateness in difficulty

In this section, the researcher finds which part of appropriateness in difficulty of the test that is mostly neglected by the English education lecturers in Parepare and how many percent of them that considered of each item. It is described in the following table.

Table 4. Frequency and Percentage of appropriateness in Difficulty Variable

Item	1		2		3		4		5		Total	
	Difficulty test		Easy test		Bloom’s Taxonomy		Test order		Empirically Difficulty test			
	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)
Always	3.33	(1)	16.7	(5)	20	(6)	30	(9)	10	(3)	16	(24)
Often	3.33	(1)	53.3	(16)	40	(12)	30	(9)	13.3	(4)	28	(42)
Sometimes	13.3	(4)	30	(9)	30	(9)	20	(6)	23.3	(7)	23.3	(35)
Seldom	53.3	(16)	0	(0)	10	(3)	13.3	(4)	33.3	(10)	22	(33)
Never	26.7	(8)	0	(0)	0	(0)	6.67	(2)	20	(6)	10.7	(16)
Total	100	(30)	100	(30)	100	(30)	100	(30)	100	(30)	100	(150)

Table 4 above shows that there are 30% (9) of lecturers that always made their test randomly between easy and difficult item. There are 53.3% (16) of lecturers that stated if their test often could be answered correctly more than 75% students. There are 30% (9) of lecturers that sometimes random the level of difficulty of the test and their tests are answered correctly by more than 75% of students. Besides, there are 53.3% (16) of lecturers whose test seldom is answered well only less than 25% of students and there are 26.7% (8) of lecturers that stated if their test never could be answered by only 25% students, commonly more of them. It means that the test that they made were easy enough because less than a quarter percent who cannot answer the test well.

Interview result between researcher and respondent strengthens the findings above. Researcher asked:

“If you made you test, did you order the easy and difficult items randomly?” Respondent 1 said:

“Yes. The important thing is that the items come up orderly as the order of teaching material. Related to easy or difficult the test is, it actually depends on the students. So, if I have tested them, I then evaluate my test, which one is easy and which one is difficult. I identified them to reconstruct in the next testing. But easy or not a test, there are three factors that possibly influence it, whether the way I was teaching, difficulty level of material or the student itself”

The tests that lecturers made in Parepare are commonly easy. It is proven by the percentage of students who can

answer the test correctly, namely 75% of all students.

Respondent 3’s statement supports it. She said:

“The test is easy for diligent students, but there are few of them always have remedial. Even they have remedial many times, the result was still same. Level of difficulty depends on students’ diligence, because we have taught them. Then when the test is difficult, should I be proud of myself? It prefers indicating that I was failed if only few of them who can answer it correctly”.

The statement above indicates whether a test difficult or not, it depends on the way students learn and prepare themselves to face examination. If they study hard by review the materials while looking for extra knowledge related to the subject, they will face the test easily. Different from those who have less preparation, they will think that the test is difficult because what lecturers have taught is same to all the students.

The test includes in an easy test, because it tests pronunciation of daily conversation (basic) to sophomore. Even though the subject, pronunciation practice II, just has been taught in that level, the conversation is done every day. This variable is neglected to consider 15.4% by the lecturers.

E. Clarity

In this section, the researcher finds which part of clear variable of the test that is mostly neglected by the English education lecturers in Parepare and how many percents of lecturers that consider for each item. It is described in the following table.

Table 5. Frequency and Percentage of clarity variable

Item	1		2		3		4		25		Total	
	Test instruction		Test duplication		Test font		Test space		Students' questions			
	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)
Always	3.33	(1)	13.3	(4)	56.7	(17)	10	(3)	0	(0)	16.7	(25)
Often	6.67	(2)	26.7	(8)	6.67	(2)	13.3	(4)	3.33	(1)	11.3	(17)
Sometimes	26.7	(8)	13.3	(4)	20	(6)	26.7	(8)	36.7	(11)	24.7	(37)
Seldom	53.3	(16)	26.7	(8)	3.33	(1)	16.7	(5)	46.7	(14)	29.3	(44)
Never	10	(3)	20	(6)	13.3	(4)	33.3	(10)	13.3	(4)	18	(27)
Total	100	(30)	100	(30)	100	(30)	100	(30)	100	(30)	100	(150)

Table 5 above shows there are 56.7% (17) of lecturers that always used font < 12 to type their test. There are 26.7% (8) of lecturers that often duplicated their test by using black and white photocopy machine even it contained pictures or diagram.

There are 36.7% (11) of lecturers that sometimes their students asked for the meaning of the questions that they made. While there are 53.3% (16) of lecturers that seldom their instruction of the test was asked for clarification by the students and there are 33.3% (10) of lecturers that never used space “one” to type their test . It means that they considered a readable test for test takers by using wider space.

In this research some lecturers sometimes use font < 12 to type their test. However, in fact, it does not mean that the test cannot be readable because it also

depends on kinds of letter as Respondent 1 said.

“Sometimes. It depends on kinds of letter because if we type by using “Maiandra” with font 11, it is the same as “Calibri” font 12. The most important is that the test is readable”.

The statement above describes that those kinds of letter and the size of font influence a lever of clarity of a text based on readable degree.

F. Appropriateness in time

In this section, the researcher finds which part of appropriateness in time of the test that is mostly neglected by the English education lecturers in Parepare and how many percent of them who considered for each item. It is described in the following table.

Table 6. Frequency and Percentage of appropriateness in Time Variable

Item	1	2	3	4	5	Total
	The number of tst takers	Relevance of test	Time allocation for each item	Time allocation best on kinds of the	Giving extra time	

	test											
	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)
Always	6.7	(2)	0	(0)	0	(0)	3.33	(1)	0	(0)	2	(3)
Often	10	(3)	16.7	(5)	0	(0)	13.3	(4)	6.67	(2)	9.33	(14)
Sometimes	10	(3)	16.7	(5)	6.7	(2)	26.7	(8)	50	(15)	22	(33)
Seldom	13.3	(4)	36.7	(11)	40	(12)	33.3	(10)	20	(6)	28.7	(43)
Never	60	(18)	30	(9)	53	(16)	23.3	(7)	23.3	(7)	38	(57)
Total	100	(30)	100	(30)	100	(30)	100	(30)	100	(30)	100	(150)

Table 6 above shows that there are 6.7% (2) of lecturers that always used oral test to test more than 30 students at once. There are 16.7% (5) of lecturers that stated if they often made few number of test, they gave longer time to students to answer it. There are 50% (15) lecturers that sometimes giving extra time for the students to answer their test from the fixed time. In addition, there are 40% (12) of lecturers that seldom neglect to put allocated time for the test that they construct and there are 60% (18) of lecturers that never used oral test to test more than 30 students at once. It means that they preferred doing oral test to few students.

Respondent 1 strengthens the findings above with his statement. He said:

“Yes, we have provided time to test. If we stop the test then continue it in the future days, it means that there will be extra time for them who take the test lately. To make it fair, we finish it at once even there

are more than thirty students”

In relation to allocated time for the test, he also added:

“Yes, whatever the topic is, I always give them chance to express their idea, mind, or opinion. The much more they said, the more improved they are”

Those statements above show that for oral test, even there are more than 30 students will take the test, It should be done once to prevent unfairness. All in all, This variable is neglected to consider 10.2% by the lecturers of English education departments in Parepare .

G. Transparency

In this section, the researcher finds which part of transparency of the test that is mostly neglected by the English education lecturers in Parepare and how many percents of the lecturers that consider for each item. It is described in the following table

Table 7. Frequency and Percentage of transparency Variable

Item	1	2	3	4	5	Total
	Good handwriting	Longer answer	Scoring based on memory	The same test	Weight score	

	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)
Always	0	(0)	3.33	(1)	0	(0)	3.33	(1)	20	(6)	5.33	(8)
Often	6.67	(2)	3.33	(1)	3.33	(1)	30	(9)	23.3	(7)	13.3	(20)
Sometimes	16.7	(5)	16.7	(5)	3.33	(1)	23.3	(7)	30	(9)	18	(27)
Seldom	13.3	(4)	23.3	(7)	30	(9)	23.3	(7)	16.7	(5)	21.3	(32)
Never	63.3	(19)	53.3	(16)	63.3	(19)	20	(6)	10	(3)	42	(63)
Total	100	(30)	100	(30)	100	(30)	100	(30)	100	(30)	100	(150)

Table 7 above shows that there are 20% (6) of lecturers that always did not put weight score to each item of the test. There are 30% (9) of lecturers that often used their previous test to test different students in the same institution. There are 30% (9) of lecturers that sometimes did not put weight score to each item of the test. There are 23.3% (7) of lecturers that seldom gave good grade to the students who had longer answer, vice versa and used the same test for different students from the same institution. There are 63.3% (19) of lecturers that never gave higher score for those whose writing were good and vice versa. It means that there were no “Hello effect” in their test.

Weight score in each item of the test offers transparent system for scoring information to the students. It aims to inform the students that which items that have higher point than others. The findings above that regarding it are supported by the statements below:

Respondent 3 said:

“The most important that I have told the scoring system orally”

The same as respondent 3, respondent 1 also said:

“Before test, I have told it orally. So they know it”

Commonly lecturers decides to announce it than put it into their test. One of example of the test which does not

have scoring information for each item can be seen in appendix 5.

Besides, some lecturers used the same test for different students in the same university. For instance who used card to test, they assumed that using card for test has lack of potency for the test to be bias because student only knows one test from one card. Here is one of respondent’s statements. Respondent 3 said:

“This kind of test (card) is not kind of test that photocopy able in which students can look for the answer and write it down to cheat. Moreover, each card has different topic. So lack of possibility for them to know all the topics in the card”

To looking at an example of kind of “Card” test, please see appendix 6. In addition, Interview result between researcher and respondent 2 supports the findings above.

Researcher asked:

“What is your reason, sir (using this test method (group video recording) every year in all classes that you teach)? Then is there any possibility for this test to be leakage?”

Respondent 2 said:

“This kind of test is more up to date. Leakage for theme might be possible, but it is impossible if what they said 100% will be same. Possibility of leakage between oral and written test must be different”

Besides, the positive thing is that “Hello effect” does not influence them in giving score to the students. Even the students have a good hand-writing, longer answer or explanation related to the questions, they keep concerning on the main focus of the test. It is proven by 63.7% of lecturers said in the questionnaire. All in all, this variable is neglected to consider 10.7% by the

lecturers.

H. Economy

In this section, the researcher finds which part of economical variable of the test that is mostly neglected by the English education lecturers in Parepare and how many percentase of lecturers who consider for each item. It is described in the following table.

Table 8. Frequency and Percentage of Economic Variable

Item	1		2		3		4		5		Total	
	Stationary in testing		Personnel involved testing		Assistance in assessing		Statistical software needs		Question file			
	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)	%	(n)
Always	0	(0)	0	(0)	0	(0)	6.67	(2)	0	(0)	1.33	(2)
Often	10	(3)	3.33	(1)	13.3	(4)	13.3	(4)	0	(0)	8	(12)
Sometimes	16.7	(5)	16.7	(5)	13.3	(4)	13.3	(4)	0	(0)	12	(18)
Seldom	40	(12)	30	(9)	26.7	(8)	33.3	(10)	16.7	(5)	29.3	(44)
Never	33.3	(10)	50	(15)	46.7	(14)	33.3	(10)	83.3	(25)	49.3	(74)
Total	100	(30)	100	(30)	100	(30)	100	(30)	100	(30)	100	(150)

Table 8 above shows that there are 6.67% (2) of lecturers that always need statistical software to interpret the result of the test. There are 13.3% (4) of lecturers that often need statistical software to interpret the result of the test. There are 16.7% (5) lecturers who sometimes use too many stationary in performing test. There are 40% (12) of lecturers who also seldom used many stationary while there are 83.3% (25) of lecturers that never did not save their own test in their files or “Bank”. It means that they kept their test in their file every time they gave them to students. Those findings above indicate that English education lecturers considered economical

side of the test. It can be seen from the percentage of “Never” level is higher than “Always” and “Often” level. Besides, the statements of two respondents support them as follow:

Respondent 1 said:

“I did everything by myself. Starting from scoring until input it to SISFO (Campus Information System), I did it by myself”.

When the researcher asked the same question the respondent 3, she also answered:

Yes, I am by my self

Above all, this variable is neglected to consider 8.9% by the lecturers. It

means that this variable is considered enough.

Among eight variables (validity, reliability, comprehensiveness, appropriateness in difficulty, clarity, appropriateness in time, transparency, and economy) that lecturers should consider in developing their own test, based on data analysis through percentage above, here are the sequence of variables which are mostly neglected by the English Education lecturers in two universities in Parepare: (1) Reliability, (2) Appropriateness in difficulty, (3) Clarity, (4) Comprehensiveness, (5) Validity, (6) Transparency, (7) Appropriateness in time, (8) Economy. The researcher finds that first four variables above are the most neglected variables, while the last four variables are the considered enough variables because there are some possibilities of other considerations which are taken by the lecturers based on the statements in interview sections.

3. The ways lecturers develop their own test

Through interview sections, every lecturer has their own way in developing their own test. Brennan in Suryabrata (1997) stated the common steps in developing a test. They are determining purpose of the test, writing test specification, writing questions, reviewing the questions, trying out the test including analysis, constructing the test, administering the test, scoring, reporting the result of the test, and using the result of the test.

Lecturers generally follow some of the steps orderly, but they have other ways which are various. In this research, the lecturers develop their test not only based on the steps above, but also they include the

way they consider a test method to apply. Respondent 1 prefers to a challenging test in which when the lecturer finds the students' weakness, he will challenge the students by use test method that can encourage the students' weakness. For instance, if the students are weak in oral performance, he will apply oral test for them. It aims to make the students learn more about it. Here is the supporting statement from respondent 1.

“Actually, it depends on what class that I face. For freshman and sophomore level, I use oral test for them. It aims to train them speak more. Different from it, for junior and senior level, I use written test because they have been good at speaking. I prefer using kinds of test in which I found students' weakness. It purposes to challenge them to improve and motivate them to master their weakness”.

Different from respondent 1, respondent 2 prefers the way they develop their test through creative test. a creative test here is a test which integrate many skills that support each other. The lecturer includes other aspects as a part of the real test. Respondent 2 is a speaking lecturer. In developing his test, he considers creativity of the test. he asks his students to make a video in which the students record themselves speaking English and asks them to make in creative whether its layout, its contain, or theme background. Here is the statement that supports the finding from respondent2:

“Yes, there are many factors that I consider in Speaking test. I asked the students to make a video in group in which they make a conversation among them related themes that I had provided. Automatically, their creativity in making

video, their expression in speaking, etc are being my consideration”

Another various way that lecturers consider in developing their own test is that a spontaneous test. The lecturer thinks that a good test is a test which can examine students' spontaneity. When a student can answer a test correctly by longer preparation, it is not something awesome for her, but when a student can answer a test spontaneously without thinking too long, it means that the student has already mastered the material comprehensively. Here is the supporting statement from respondent 3:

“A spontaneous test, it can be seen which students can answer spontaneously and orderly, it means that they can master the materials. In contrary if they think too long, it means that the materials are out of their head. Have a look to this test in which the students should answer the test spontaneously”

4. Conclusion

As the end of this research, the researcher would like to give conclusions as follow: 1) From eight variables of this

research, here is the sequence of variables that mostly lecturers neglected (1) Reliable, (2) Appropriate in difficulty, (3), Clear (4) Comprehensive, (5) Valid, (6) Transparent, (7) Appropriate in time, (8) Economical. 2) The sequence above does not represent broadly lecturers' consideration in developing their own test. For those the first four variables, they became the most neglected factors among them, but for the last four variables, they might be considered by the lecturers. However based on the interview to three respondents, each respondent showed their own special consideration when they develop their own test.

In relation to the conclusion above, the researcher gives some suggestions for language test development as follow: 1) The institution should facilitate the lecturers a seminar or workshop in certain period (whether once in a semester or once in a year) to update the quality of the instrument of evaluation to prevent the neglects of considerations in developing a test 2) It is good to make lecturer cognate in each institution to share various innovations of developing language testing .

References

Aiken, Lewis R. (1971). *Psychological Testing and Measurement*. Allyn and Bacon. Inc

Anderson, Scarvia B. et al. (1976). *Encyclopedia of Educational Evaluation*. Sanfransisco, Washington, London: Jossey Bass Publishers

Anyonymous. (2010). *Undang-Undang*

Sisdiknas (Sistem Pendidikan Nasional). Jakarta: Fokus Media

Arikunto, Suharsimi.(2005). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara

Astuti, Feny Andri.(2013). *The Analysis of Teacher-Made English Mid-Term Test for Twelfth Grade Students of the Odd Semester 2012/2013*

- of SMK N 1 Kudus. Unpublished Theses. Muria Kudus University
- Aswad, Muhammad. 2017. "The Effectiveness English Camp (A Model in Learning English as the Second Language)." In *ASEAN/Asian Academic Society International Conference Proceeding Series*.
- Biggas, John B. J. More, Philip. (1993). *The Process of Learning. Third Edition*. Australia: Macarthur Press Books
- Bruce W. Tuckman .*Using Frequent Testing to Increase Students' Motivation to Achieve* .The Ohio State University Copyright 2000-2008 The Ohio State University. All Rights Reserved.
- Burke, Kay. (1999). *How to Assess Authentic Learning*, 3rd Edition. SkyLight Training and Publishing Inc. Reprinted by permission of Pearson SkyLight, www.skylightedu.com
- Cross. A. (1973). *Home Economics Evaluation*. Colombus Ohio: A Bell and Honel Company
- Daggett, Willard R. Cobble, Jeffrey E. Gertel, Steven J. (2008). *Color in An Optimum Learning Environment*. International Center for Leadership in Education. Pdf
- Gay. L.R. (2006). *Educational Research: Competencies for Analysis and Applications*. Ohio: Pearson Merrill
- Prentice Hall
- Heaton, J.B. (1998). *Writing English Language Test*. New York: Longman
- Henning, Grant. (1987). *A Guide to Language Testing*. Boston: Heile & Heinle Publisher.
- Huges ,Arthur. (1989). *Testing for Language Teacher*. New York: Cambridge University Press.
- Jabu, Baso. (2007). *English Language Testing*. Makassar: UNM
- Junaedi, Iwan. (2010). *Penulisan Prediksi Soal UN Matematika SMA*. A Paper presented in TOT Guru Mata Pelajaran yang diUN-kan by Dinas Pendidikan Provinsi Jawa Tengah
- McNamara, Tim. (2000). *Language Testing*. NY: Oxford University Press
- Miles, M.B & Huberman, A.M. (1984). *Qualitative Data Analysis*, 16 Newbury Park, CA: Sage
- P. Harris, David. *Testing English as a Second Language*. Bombay. Tata McGraw- Hill
- Purwanto, Ngalim. (2002). *Prinsip-Prinsip dan Teknik Evaluasi Pengajaran*. Bandung: Remaja rosda karya
- Ramadhan, Mohammed. (2011). *Characteristics of a good test*. <http://eltguide.wordpress.com/2011/1>

[2/28/12-characteristics-of-a-good-test/](#). Accessed on March 7th, 2014

- Rea-Dickins, et all. (1992). *Evaluation*. NY: Oxford University Press.
- Richards, J.C & Rodgers, T.S. (2001). *Approaches and Methods in Language Teaching*. In Savage, K.L,
- S. Madsen, Harol. (1983). *Technique in Testing*. NY: Oxford University Press.
- Sari, Octaviani, Mayustari. (2010). *Item Analysis of English Formative test Made by English Teacher SMA 1 IV Angkek*. English Lecturer of English Department of STKIP PGRI Sumatra Barat.
- Slameto . (1999). *Evaluasi Pendidikan*. Jakarta. Bumi Aksara
- _____. (2010). *Belajar dan Faktor-Faktor yang Mempengaruhinya*. Jakarta:Rineka Cipta
- Sudjana, Nana. (2008). *Penilaian Hasil Proses Belajar Mengajar*. Bandung: Remaja Rosda Karya
- Sukardi. (2009). *Evaluasi Pendidikan. Prinsip & Operasionalnya*. Jakarta: Bumi Aksara
- Suryabrata, Sumadi. (1984). *Pembimbing ke Psikodiagnostik*. Yogyakarta: Andi Offset
- _____. (1995). *Psikologi Pendidikan*. Jakarta: Raja Grafindo Persada
- _____. (1997). *Pengembangan Tes Hasil Belajar*. Jakarta. Raja Grafindo Persada
- Thoha, Chabib. (2003). *Teknik Evaluasi Pendidikan*. Jakarta: Raja Grafindo Persada
- Widowati, Diah Retno. (2011). *Item Analysis of a Teacher-Made Try-Out Test of UAN 2010/2011 of Junior High Schools in Malang*. Unpublished Thesis, English Language Teaching, Faculty of Letters, State University of Malang.
- Widyoko, S. Eko Putra. (2009). *Evaluasi Program Pembelajaran*. Yogyakarta: Pustaka Pelajar.
- Wringstone-Justman-Robbin. (1956). *Evaluation in Modern Education*. NY: American Book Company