

TEACHER-MADE TEST IN EFL LISTENING COMPREHENSION

Yazid Rukmayadi (1202023)

Indonesia University of Education

Email: yazidrukmayadi@yahoo.com

ABSTRACT

This research is aimed at assessing teacher's made test of listening comprehension in the term of construct validity and passage characteristics (type, length, speed of speech, and accents).

Descriptive method through quantitative-qualitative approach is employed in this research. The research is conducted on 2nd grade students of one of senior high school in Bandung as population, and 30 students are taken as sample. ANA-TEST is employed to find out the construct validity of the test, and document analysis is conducted to investigate the characteristics of the passages in listening test items. The finding shows that from 20 test items being analyzed **18 items** were valid and **2 items** (No. 6 & 15) were considered invalid. The finding also showed that the reliability of the instrument measured was **0.59**. The data from document analysis showed that the passage consisted of monologue and dialogue (conversation), the length was 14 minutes, 18 seconds. The average of speed of speech was 120 wpm for directions or monologues, and 150 wpm for conversation with non-regional accent. Finally, from document analysis had also been revealed that the accent was non-regional since the speakers were non-native speakers.

Finally, it is recommended for the teacher to have more attention on construct validity of the test while developing a test, and also give more attention on passage characteristics in order to meet a good test items.

INTRODUCTION

Listening is one of the fundamental elements of comprehending a language. Harmer (2004) stated that listening is one of the receptive skills because people understand the message from what they hear. Listening is the process of thinking or changing the meaning to what one hears (Tompkins and Hoskisson (1991:108)). Furthermore, listening is integrated with the other three basic skills: speaking, reading, and writing. In aural communication there are two skills involved, they are listening and speaking and also in language learning. According to Brown (2001:234) language learners learn to what to speak from what they listen. A student who learns how to pronounce a certain word is surely given example by the teacher in order to pronounce correctly.

In spite of its importance, EFL learners often regard listening as the most difficult language skill to learn (Hasan, 2000; Graham, 2003). As Vandergrift (2007) points out, one of the reasons might be that learners are not taught how to learn listening effectively. A narrow focus on the correct answer to comprehension questions that are often given in a lesson does little to help learners understand and control the process leading to comprehension. When learners listen to spoken English, they need to perceive and segment the incoming stream of speech in order to make sense of it. The listener cannot refer back to the text in contrast to a reader who usually has the opportunity to refer back to clarify understanding.

Furthermore, the problems dealing with listening also arise in the term of testing process. Listening has traditionally been the forgotten skill in the term testing (Douglas, 1988). Buck (1991: 67) attributes this neglect to the lack of a widely-accepted theory of listening comprehension, and goes on to state, "It seems that in practice test constructors are obliged to follow their instincts and just do the best they can when constructing tests of listening comprehension". Obviously, this haphazard approach to testing listening

presents serious implications for the validity of these assessments. Fortunately, in the last decade the assessment of listening has attracted increasing amounts of attention, and a great amount of research has been conducted on the subject.

A great amount of research has been conducted in testing listening. For example; research by Buck, 1991, 2001; Buck & Tatsuoka, 1998; Dunkel, Henning, & Chaudron, 1993; Richards, 1983; Rubin, 1994 have described the necessity of defining the concept of listening comprehension test, even though an appropriate definition is still elusive, and there seems to be a general statement that there is no widely-accepted definition (Bejar, Douglas, Jamieson, Nissan, & Turner, 2000; Brindley, 1998; Buck 1994, 2001). Part of the problem lies in the fact that because so many different processes and aspects are involved in EFL listening comprehension, providing a global, comprehensive definition may be impossible. Richards (1983) explains how EFL listening varies according to the purpose of the learners such as listening for social interaction, information, academic listening, listening for pleasure, or for some other reason.

Nowadays, multiple choices are commonly used to measure students' achievement, especially in the listening assessment. Alderson (2000:211) argues that this form of test is well-liked because it can be used to control the range of students' answers as well as to control the variety of students' answers. Another reason why this form of test is commonly used is because the result can be checked with the computer so that it saves time, money, and energy. This test is also considered as the most objective form of test.

This study aims at finding the characteristics of teachers-made test in English final term test of 2nd grade senior high school students in Bandung. Document analysis and statistical computation (with ANA-Test) are employed to find out passage characteristics and construct validity of the test. Hughes (2003) states that analyzing construct validity of the tests is important to be employed in order to know whether the test is good or not.

In addition, Hughes (2003: 162) also states that texts should be specified as fully as possible. Here are the characteristics of the passage which influence in the process of testing listening comprehension according to Hughes (2003):

- **Text Type** might be first specified as monologue, dialogue, or multi-participant, and further specified: conversation, announcement, talk or lecture, instructions, directions, etc.
- **Length** may be expressed in seconds or minutes. The extents of short utterances or exchanges may be specified in terms of the number of turns taken.
- **Speed of speech** may be expressed as words per minute (wpm) or syllables per second (sps).
- **Accent** may be regional or non-regional.

METHODOLOGY

This research employs descriptive method through quantitative-qualitative approach. This research used 30 samples of students' answer sheets from the 2nd grade senior high school English final semester test in Bandung. The research used some instruments to find out the data, the instruments are:

a. Statistical Computation (ANA-Test)

ANA-TEST is employed to quantitatively analyze the construct validity of the test items (validity, reliability, and difficulty).

b. Document analysis

Document analysis is employed to analyze the characteristics of the passages in the term of *text type*, *length*, *speed of speech*, and *accents*. *Text Type* might be first specified as monologue, dialogue, or multi-participant, and further specified: conversation, announcement, talk or lecture, instructions, directions, etc. *Length* may be expressed in seconds or minutes. The extents of short utterances or exchanges may be specified in terms of the number of turns taken. *Accent* may be regional or non-regional. Finally, *Speed*

of speech may be expressed as words per minute (wpm) or syllables per second (sps). Reported average speeds for samples of British English are:

Text types	wpm	sps
Radio monologues	160	4.17
Conversations	210	4.33
interviews	190	4.17
Lectures to non-native speakers	140	3.17

(Tauroza and Allison, 1990, cited by hughes, 2003)

FINDING AND DISCUSSION

A. Construct Validity

- *Test Validity*

Validity is a matter of degree to extend the result of study as one way to measure the validity through carrying out item of instrument analysis (Hatch and Farhady, 1982: 251). According to Beanland, et.al, (1999) validity of research instrument is the degree to which the instrument measures what it is supposed to measure. Validity is closely related to reliability because for an instrument to be valid, it must be reliable (beanland et.al, 1999). It is also important to remember that instrument may in fact be reliable even when they are not valid (Beanland et.al, 1999, Polit & Hungler, 1999).

Commonly assessing test validity employs Pearson product moment correlation. The formula, as follows:

Note:
$$r_{xy} = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{\{N \sum X^2 - (\sum X)^2\} \{N \sum Y^2 - (\sum Y)^2\}}}$$

r_{xy} : coefficient correlation between variable X and Y

X : item which its validity is assessed

Y : total score gained by the sample

(Arikunto, 2003: 146)

Even though, in this study ANA-Test was applied to measure validity with Pearson product moment correlation type. Here is the result of the analysis:

The Validity of Each Item

Items	Value	Interpretation
1	0.415	Valid
2	0.684	Valid
3	0.127	Valid
4	0.279	Valid
5	0.348	Valid
6	NAN	Invalid
7	0.228	Valid
8	0.060	Valid
9	0.496	Valid
10	0.080	Valid
11	0.240	Valid
12	0.505	Valid
13	0.477	Valid
14	0.374	Valid
15	NAN	Invalid
16	0.268	Valid
17	0.149	Valid
18	0.342	Valid
19	0.238	Valid
20	0.216	Valid

Based on the table above (table 3.2), there were 18 items valid. Then, those appropriately became the instrument to apply in testing listening. The rest of 2 items (6 &15) were invalid, so those were considered not appropriate to use as the instrument of the test.

Test Reliability

Reliability is the extent to which the result can be regarded consistent or stable (Brown, 1990: 98). According to Beanland et.al (1999) reliability is the degree to which an instrument produces the same results with repeated administration. A high level of reliability is particularly

important when the effect of an intervention on knowledge is measured using a pre-test/post test design.

To interpret the coefficient of reliability, the following criteria are employed:

Table 3.4
The Criteria of Reliability Test

Coefficient Reliability	Interpretation
0.00 - 0.19	Very Poor
0.20 - 0.39	Poor
0.40 - 0.59	Moderate
0.60 - 0.79	Good
0.80 - 1.00	Excellent

(Sugiono, 2001: 149)

In this study, ANATEST was applied to reveal the reliability of instrument. The result showed that the reliability of the instrument measured was **0.59**. In keeping with Sugiono (2001: 149), the value of alpha is considered **moderate** for the items. Thus, the items were appropriate enough to be the instrument given to learners in the study.

Test difficulty

Another requirement that needs to be considered as excellent instrument is difficulty test. Arikunto (1993: 209) argued that difficulty test aims to get the level of difficulty for each item of the instrument.

Based on the results analyzed by ANA-test, 4 items (3, 6, 13, 16) were categorized very difficult. 3 items (5, 10, 19) were categorized difficult. Meanwhile, 5 items (4, 9, 12, 18, 20) were categorized moderate. 4 items (1, 2, 7, 14) were considered easy. The rest of 4 items (8, 11, 15, 17) were categorized very easy.

Characteristics of the Passages

- ***Text Type***

Text type according to Hughes (2003) might be first specified as monologue, dialogue, or multi-participant, and further specified: conversation, announcement, talk or lecture, instructions, directions, etc. The result revealed from document analysis showed that the passage consisted of two types of text, monolog and dialog. Monolog was in the term of radio monologues or direction, and dialog was in the term of conversation between two or three persons. The questions of the test were based on the conversation presented by listening audio.

- **Length**

Length as stated by Hughes (2003) may be expressed in seconds or minutes. The extents of short utterances or exchanges may be specified in terms of the number of turns taken. Based on document analysis, the length of listening texts presented by listening audio was 14 minutes and 18 seconds (14:18) for 20 test items. It means that the average length for one item test needed about one minute or less.

- **Speed of speech**

Hughes (2003) stated that speed of speech may be expressed as words per minute (wpm) or syllables per second (sps). Based on documents analysis through listening audio, the speed of speech was 120 wpm for the average of radio monologues or directions, and 150 wpm for average of conversations. Differs to the speed of speech standard of British English (Hughes, 2003), the speed was slower than the British one.

- **Accent**

According to Hughes (2003), Accent may be regional or non-regional. Based on document analysis, the accent was indentified as non-regional accent due to the speakers are non-native speakers. But in average, each speaker had good American accent.

CONCLUSION

This study aims at finding the characteristics of teachers-made test in English final term test of 2nd grade senior high school students in Bandung. Document analysis and statistical computation (with ANA-Test) are employed to find out passage characteristics and construct validity of the test. Hughes (2003) states that analyzing construct validity of the tests is important to be employed in order to know whether the test is good or not. The finding showed that from 20 test items being analyzed **18 items** were valid and **2 items** (No. 6 & 15) were considered invalid. The finding also showed that the reliability of the instrument measured was **0.59**. In keeping with Sugiono (2001: 149), the value of alpha is considered **moderate** for the items.

Underpinned by Hughes's theory of passage characteristics (Hughes, 2003), the data from document analysis showed that the passage consisted of monologue and dialogue (conversation), the length was 14 minutes, 18 seconds. The average of speed of speech was 120 wpm for directions or monologues, and 150 wpm for conversation. Finally, from document analysis had also been revealed that the accent was non-regional since the speakers were non-native speakers.

Finally, it is recommended for the teacher to have more attention on construct validity of the test while developing a test, and also give more attention on passage characteristics in order to meet a good test items.

REFERENCES

- Arikunto, S. (2003). *Prosedur Penelitian: Suatu Pendekatan Praktek. Edisi Revisi V*. Jakarta: Rineka Cipta.
- Bejar, I., Douglas, D., Jamieson, J., Nissan, S., & Turner, J. (2000). *TOEFL 2000 listening framework: A working paper (TOEFL Monograph Series Report No. 19)*. Princeton, NJ: Educational Testing Service.
- Brindley, G. (1998). Assessing listening abilities. *Annual Review of Applied Linguistics*, 18, 171-191.
- Brown, H.D. (2001). *Teaching by Principles: An Interactive Approach to Language Pedagogy (second Edition)*. White Plains: Addison Wesley Longman, Inc.
- Buck, G. (1991). The testing of listening comprehension: An introspective study. *Language Testing*, 8, 67-91.
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.
- Buck, G., & Tatsuoka, K. (1998). Application of the rule-space procedure to language testing: Examining attributes of a free response listening test. *Language Testing*, 15, 119-157.
- Douglas, D. (1988). Testing listening comprehension in the context of the ACTFL proficiency guidelines. *Studies in Second Language Acquisition*, 10, 345-61.
- Graham, S. (2003). Learner strategies and advanced level listening comprehension. *Language Learning Journal*, 28, 64-69. doi: 10.1080/09571730385200221
- Harmer, J. (2004). *The Practice of English Language Teaching*. Cambridge: Longman
- Hasan, A. (2000). Learners' perceptions of listening comprehension problems. *Language, Culture and Curriculum*, 13, 137-153. doi: 10.1080/07908310008666595
- Hughes, A. (2003). *Testing for Language Teachers (second edition)*. Cambridge (UK): Cambridge University Press.

- Richards, J. (1983). Listening comprehension: approach, design, procedure. *TESOL Quarterly*, 17, 219-40.
- Rubin, A. (1980). *Theoretical taxonomy of the difference between oral and written language*. In R. Spiro, B. Bruce, & W. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 411-438). Hillsdale, NJ: Erlbaum.
- Rubin, J. (1994). A review of second language listening comprehension research. *Modern Language Journal*, 78, 199-221.
- Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching*, 40, 191-210. doi: 10.1017/S0261444807004338