AN ANALYSIS OF A MIDTERM ENGLISH TEST
OF THE SEVENTH GRADE ACCELERATED CLASS
AT SMPN 3 MALANG

Arin Eka Primadani, Gunadi Harry Sulistyо
Email: arina.gee2@gmail.com; emailgun.now@gmail.com
State University of Malang

ABSTRACT: This study was conducted to evaluate the quality of the teacher-made test of the midterm English test of the 7th grade accelerated class at SMPN 3 Malang. Using the qualitative and quantitative method, the quality of the test was analyzed based on the content validity, test reliability, item discrimination, item validity, and the effectiveness of the distracters. The results of the analysis shows that the quality of the test is low in terms of the difficulty level, item discrimination, item validity, and the distracters even though the test reliability is acceptable and the items are 100% valid in covering the materials presented in the curriculum.

Keywords: item analysis, midterm English test, 7th grade accelerated class

Test is an important part of teaching learning process that cannot be separated in the implementation of the teaching and learning process itself. Djiwandono (2008:1) states that in language learning, learning is implemented to obtain several learning objectives that are already identified and formulated based on thorough interpretation of the needs that have to be obtained. Both teaching and testing are so closely interrelated that it is virtually impossible to work in either field without being constantly concerned with the other (Heaton, 1998:5). The objectives of learning can be evaluated by having a test, since the test is constructed to find out the achievement of the learners in teaching and learning process. Then, from the result of the test, it can be determined if the learning objectives are obtained successfully. Regarding that matter, it is obvious that conducting a test cannot be ignored in the implementation of teaching and learning process.

The result of the test is used to evaluate the progress of teaching and learning process, since the performances of the learners have to be evaluated progressively. Besides, the result of the test can be used to see how successful the teaching and learning process is implemented. As stated by Woods (2005:25), this is important as testing can often influence the nature of what is taught and how it is taught. It is quite clear then that by having a test, the teacher can see how far the learners understand the materials that he or she delivered and how well the materials are delivered. Therefore, the good test will be able to provide the accurate information about learners’ performance and the quality of the teaching and learning process.
The achievement of the learners toward the instructional objective of teaching learning can be brought up by conducting a test. Brown (1994:385) points out that a test measures a person’s ability or knowledge. Based on that theory, by having a test, the teacher is able to find out the background knowledge of the learners and the ability of the learners in understanding the materials. Another language testing objective carried out by Harris (1969:3) is to measure the extent of student achievement of the instructional goals. It means that a test in language learning is as a measurement of the progressive achievement that is already obtained by the students in a certain period of time within the learning program. Thus, by knowing the students’ achievement from the test results, it can be determined if a group of students are ready for entering a higher level of learning program.

As the measurement to evaluate the students’ performance in the teaching and learning process, the test should meet the criteria of a good test. There are several aspects that constitute the criteria of a good test. According to Weir as cited in Woods (2005), the concern is how far we can depend on the results that a test produces or, in other words, could the test produce results consistently. It is clear then that the test which can give consistent results is considered reliable. It is supported by Woods (2005:25) who states that the underpinning test design really attempts to have high validity whilst retaining reliability. It can be said that validity and reliability cannot be ignored when the teacher is constructing a test. Additionally, Sulistyo (2007:21) states that a good test must meet the requirements: reliability, validity, practicality/usability, and economy. If a test does not meet the requirements of a good test, the test will produce biased scores that will not reflect the real ability of the test takers. Then, it is obvious that the good test must meet those criteria so that the results of the test can be definitely relied on.

The curriculum that is implemented in Indonesia when this study is conducted is *Kurikulum Tingkat SatuanPendidikan* that lets the school develop their own teaching and learning materials based on standard of content. This policy gives the school opportunities to develop their own teaching and learning materials and the test as the evaluation device. The schools can develop their own teaching and learning materials based on strengths or weaknesses and the social context of each school. The test that is developed in this context is known as teacher-made test. The teacher-made test here is constructed and administered by the teacher. Hence, the items of the test are not analyzed systematically by the teacher. Arikunto (2003:147) states that the teacher-made test is constructed from the items that are
commonly not tried out, analyzed, and revised first. Therefore, based on that case, the quality of the teacher-made test is questionable.

Regarding the quality of the teacher-made test, it is obvious that the teacher-made test needs to be tried out and analyzed. The results of the analysis will show which items have good quality and which items need to be revised. In examining the quality of the teacher-made test, the content validity, reliability, item difficulty, item discrimination, item validity, and the effectiveness of the distracters need to be analyzed as well. The validity and reliability are examined on the test as a whole, while the item difficulty, item discrimination, item validity, and the effectiveness of the distracters are used to analyze the quality of the test items.

A similar study was conducted by Ratnafuri (2011). She analyzed content validity, reliability, practicality, item difficulty, item discrimination, and the effectiveness of the distracters of the teacher-made English Test in the final test of the 2nd semester of 2010/2011 of the first year students of SMPN 1 Kauman, Tulungagung. The results of this study show that the test has moderate content validity, good reliability on the multiple choice items, and the practicality is good. Only 18% of the test items are good in terms of item difficulty, and the index discrimination is considered to be low. The last, there are many distracters which are ineffective.

Based on the previous description, a study focusing on the analysis of a test in SMPN 3 Malang needs to be conducted. The data of Dinas Pendidikan in Malang (Rekapitulasi Statistik PPDB untuk Sekolah Menengah Pertama, online) shows that SMPN 3 Malang has the highest input of 9,917 compared to other schools. Moreover, SMPN 3 Malang is an international standard school when this study was conducted. It can be concluded that SMPN 3 Malang is considered as a good school in Malang. For that reason, it is believed that the teachers there also have good quality in conducting teaching and learning. Thus, the test that is constructed and administered by the teachers in the school should have good quality in measuring students’ performance.

Realizing the need for good quality teacher-made test, the quality of the teacher-made test needs to be analyzed. Based on an informal interview with the English teacher, it is summarized that the teacher did not have enough time to try out, analyze, and revise the test items. Thus, the teacher just constructed the test based the blueprint and standard of content only.
Based on the result of the data analysis, the midterm English test of 7th grade accelerated class at SMPN 3 Malang has very good content validity as it reflects 100% of the test materials and materials stated in curriculum. The reliability of the test is considered to be acceptable for a classroom test with a coefficient of .531 estimated from the first re-administration data and .783 estimated from the second re-administration data. The difficulty level of the test is considered to be low since the test items are mostly easy for the students. Only 40% of the test items have good discrimination index estimated from the first re-administration data analysis, and 48% of them have good discrimination index estimated from the second re-administration data. Hence, most of the test items cannot show the equality of the score of each item with the total score since based on the first re-administration data analysis only 26% of the items have good item validity index, and only 24% of the items have good item validity based on the second re-administration data analysis. Lastly, from the data analysis of the first and second re-administrations, only seven items (14%) have all their distracters work properly. In contrast, there is no logical relationship between the result of the content validity analysis and the item analysis. The content validity analysis result shows that 100% of the test items are valid in terms of the content, but the item analysis results show that most of the items are not good. This is considered to be the weakness of this study.

METHOD

The design used in this study is descriptive evaluative because this study is aimed to evaluate the quality of the teacher-made test of midterm test of the 2nd semester academic year 2012/2013 of 7th grade accelerated class at SMPN 3 Malang. This study was conducted using a qualitative and quantitative analysis method. The qualitative analysis was used to evaluate the content validity of the test. The quantitative analysis was used to evaluate reliability, item difficulty, item discrimination, item validity, and the effectiveness of the distracters.

The data sources for this study were the school documents and students’ score. The school documents were the test and the blueprint made by the teacher to develop the test. The test consists of fifty items in the form of multiple-choice format. The materials presented in the test were short functional texts and short monolog essays. The short functional texts are in the form of notice, caution, greeting card, announcement, letter, advertisement, invitation, and e-mail. The short monolog essay is a descriptive text.
The students’ scores were obtained from the re-administrations of the test in two different schools because the raw data of students’ scores of SMPN 3 Malang no longer existed. Moreover, the re-administration of the test was aimed to see whether the test is reliable for the different group of students of the same level. The students as the test takers in the re-administration of the test were the students who are in 7th grade accelerated program of SMPN 1 Malang and SMPN 5 Malang. There were nineteen students of SMPN 1 Malang and seventeen students of SMPN 5 Malang who become the test takers in the re-administration of the test. The students were in the academic year 2012/2013. The scores of these students were the data to analyze the reliability, item difficulty, item discrimination, item validity, and the effectiveness of distracters of the test.

A table of specification was used to analyze the content validity of the test. A Statistic method was used to analyze the reliability, item difficulty, item discrimination, item validity, and the effectiveness of the distracters. The data were analyzed using ITEMAN software version 3.00.

FINDINGS AND DISCUSSION

The results of the analysis show the content validity, reliability, item difficulty, discrimination, and effectiveness of the distracters.

Content Validity

There are fifty items of the test that were analyzed. Based on the results of the analysis, the teacher used two standards of competences which cover reading and writing skills as the bases of developing the blueprint. After analyzing the blueprint made by the teacher, the researcher found that all of the items (100%) follow the standard of competence stated in the curriculum. The test items only cover two skills, i.e. reading and writing. There are 60% of total materials which cover short functional texts, while the materials which cover the short monolog essays are 40%.

The findings of the content validity analysis show that 100% of the test items cover the materials stated in the curriculum and the test items represent the materials that should be tested to the students. The indicators constructed by the teacher also represent the materials in the curriculum that should be given to the students. These findings are in line with Arikunto’s (2003:67) statement that a test has high content validity if it measures
certain objectives which are in line with the materials that should be given. It means the test items are able to represent the materials stated in the curriculum.

**Reliability**

The reliability index is indicated by the value of alpha estimated in the analysis using the software. There are two results of reliability since the re-administration of the test is administered in two different schools. The reliability of the test from the first re-administration is .531 which means that the items in the test are good enough for a classroom test and a few items need to be improved. Based on the result of the second re-administration, the reliability coefficient is .793 which means the test is considered to have high reliability and the test items are very good for the classroom test.

Based on the results of reliability analysis from the first re-administration, the midterm English test of the 2nd semester of the academic year 2012/2013 of 7th grade accelerated class at SMPN 3 Malang have adequate reliability. Next, based on the result of the second re-administration, the test can be assumed to have a high reliability. Hence, the result of the reliability estimated from the second re-administration in this study that is .793 is in the same range with the reliability estimated in the previous study revealed by Ratnafuri (2011) which lies in .76 and is categorized as acceptable reliability for the test. Thus, based on the first and second re-administration data analysis, the reliability of the midterm English test of 2nd semester academic year 2012/2013 of 7th grade accelerated class at SMPN 3 Malang is accepted for the classroom test.

**Item Difficulty**

Based on the analysis of the item difficulty from the first re-administration, the item difficulty index indicates that that there are two items (4%) that are categorized as difficult items. Those two items are not good for the test, so those items must be revised. There are thirteen items (26%) that are categorized as moderate items. This means that those items are within the moderate range of difficulty level. Those items are good to be given to the students because the items are expected to measure students’ performance appropriately. There are thirty five items (70%) that are categorized as easy items. In the analysis of the second re-administration data, the item difficulty index indicates that there are three items (6%) that are categorized as difficult items and four items (8%) that are categorized as moderate items. This means that those items are in a moderate range of
difficulty level. Those items are good to be given to the students because the items are
good in measuring students’ performance or ability. There are forty three items (86%) that
are categorized as easy items.

The data analysis of the first re-administration data shows that the difficulty index
of the test is low since there are only thirteen items (26%) that are categorized as good
items with moderate difficulty level and based on the data analysis of the second re-
administration, the test only has four items (8%) with moderate difficulty level. Thus,
based on the first and second re-administration data analyses, the difficulty level of the
items in the midterm English test of 2nd semester academic year 2012/ 2013 of 7th grade
accelerated class at SMPN 3 Malang are considered to be low since not more than thirteen
items which have moderate difficulty level.

**Discrimination**

The analysis of first re-administration data in terms of the item discrimination of
the test shows that there are twenty seven items (54%) which poorly discriminate students’
ability. There are three items (6%) which are categorized as less discriminating items.
Those three items must be revised. There are four items (8%) with discriminating
category. Those items are accepted in the test. There are sixteen items (32%) categorized
as very good items because they have very discriminating indices. Based on the second re-
administration data analysis, there are twenty two items (44%) which poorly discriminate
the students’ ability. Those twenty two items are rejected and cannot be used in the test.
Next, there are four items (8%) which are categorized as less discriminating items and
those items must be revised. There are two items (4%) with discriminating category.
Those two items are accepted and can be used in the test. There are twenty two items
(44%) categorized as very discriminating items because they have very good
discrimination indices.

There are many items of the test which do not have an adequate discrimination
index which means those items cannot discriminate the students’ ability properly and
cannot be accepted in the test. This result is not in line with Heaton (1988:179) who states
that the discrimination index of an item indicates the extent to which the item
discriminates between the testees, separating the more able testees from less able. In short,
based on the data analysis of the first and second re-administration data, there are many
items of the midterm English test of 2nd semester academic year 2012/2013 of 7th grade accelerated class at SMPN 3 Malang which must be revised.

**Item Validity**

The item validity analysis of the test based on the first re-administration data shows that there are seven items (14%) categorized as poorly valid items and those items are rejected. There are six items (12%) categorized as less valid items. Next, there are twelve items (24%) categorized as adequately valid items. There is only one item (2%) categorized as valid item. Unfortunately, there are twenty four items (48%) which are not valid in representing the equality with the total score of the items. Then, based on the second re-administration data analysis, there are five items (10%) categorized as poorly valid items. There are thirteen items (26%) categorized as less valid items and those items must be revised. Seven items (14%) have adequate item validity and those items are acceptable. There are four valid items (8%) in the test and those items are quite good for the test. There is only one item (2%) categorized as highly valid items. Unfortunately, there are twenty items (40%) which are not valid in representing the equality with the total score of the items.

According to Arikunto (2003:75), estimation of the item validity index can show which item meets the requirement of good item. The estimation of validity index indicates that in the midterm English test of 2nd semester academic year 2012/2013 of 7th grade accelerated class at SMPN 3 Malang, only few items meet the requirements of a good item. The results of the analysis from two re-administrations reveal that more than half of the total items of the test do not have an adequate item validity index.

**Effectiveness of the Distractors**

The results of the distracters analysis from the first re-administration shows that 30% of the multiple-choice items have inadequate distracters; 32% have only one distracter that works and the other two distracters do not attract the students to choose; 24% have two distracters which are able to attract students to select; and 14% have all the distracters work well. The results of the second re-administration data analysis shows that 38% of the multiple-choice items have inadequate distracters since there is no student who chooses the whole distracters in those numbers; 26% have one distracter that works; 32% have two distracters which can attract students to select, whereas another one does not
work in distracting students’ performance; and 4% have distracters which all of them work well.

The distracters analysis of the first re-administration data shows that the distracters that work properly are only the distracters in seven test items (14%). Based on the data analysis from the second re-administration, there are only two items (4%) of the test that have all their distracters work. There are so many items which have no distracter that works. There are so many items in the test that do not have distracters which meet the requirement of effective distracters proposed by Badan Standard Nasional Pendidikan (2010), that is, a distracter works if at least 5% of the test takers choose it and the test takers who choose it are mostly from lower group who do not master the materials. To sum up, based on the first and second re-administration data analyses, many items of the midterm English test of 2nd semester academic year 2012/2013 of 7th grade accelerated class at SMPN 3 Malang have poor distracters which cannot work properly and those distracters must be revised.

The content validity of the midterm English test of 2nd semester academic year 2012/2013 of 7th grade accelerated class at SMPN 3 Malang is considered to be high. The test is considered to have high content validity because it covers all materials (100%) stated in the curriculum and the indicators also describe the objectives properly. Secondly, the reliability of the test is considered to be acceptable for a classroom test. It is considered to be acceptable since the reliability estimated from the first re-administration is .531 which means good and from the second re-administration is .793 which means adequate reliability. Third, the difficulty level of the test is considered to be low since the test items are mostly easy for the students’ and the items need to be revised or discarded. The next conclusion of this study is about the index discrimination of the test items. The discrimination level of the test is considered to be moderate since there are only about 40% of the test items which have very good discrimination index. The next point to be concluded is about the level of validity of the test items. The item validity level of the test items is considered to be low since there are only few items which have good item validity. Based on the data analysis of the re-administration 1, there is only one item which has good item validity, and based on the data of the second re-administration, there are only five items which have good item validity. Lastly, the quality of the distracters of the test items in the midterm English test of 2nd semester academic year 2012/2013 of 7th grade accelerated class at SMPN 3 Malang is considered to be low. Based on the data
analysis of the first re-administration, there are only seven items which have proper
distracters, and from the second re-administration there are only two items which have
proper distracters.

CONCLUSIONS AND SUGGESTIONS

Based on the analysis of the first and second re-administration data, the test is
considered to have low quality of the items in terms of the difficulty level, item
discrimination, item validity, and the distracters. In contrast, there is no logical
relationship between the results of content validity analysis with the item analysis of each
item. The result of content validity analysis shows that the test items are 100% valid in
covering the materials presented in the curriculum, but the results of the item analysis
show that most of the items are not good. This case appears since there is no expert
involved in this study to verify the content validity analysis.

In order to have good quality in measuring students’ achievement, the test should
meet the requirements such as, the test items must be constructed based on the standard
competence stated in the curriculum to have good content validity, the test should be
reliable, and each item should meet the requirements of a good test item. Although
evaluating the test items in order to meet the requirements of the good test is quite time
consuming, the teachers should do that evaluation as their job and in order to construct the
test which is able to measure students’ competence properly. The teachers should really
understand the concept in constructing and developing the test items, try out the test, and
analyze the test they have constructed.

It is suggested that teachers involve experts in analyzing the test especially in
analyzing the content validity. By involving the expert as the judge of the content validity
analysis, the result of the analysis will be more valid and can be relied on.

REFERENCES

Ahmann, J. S. and Glock, M. D. 1981. Evaluating Student Progress: Principles of Test
and Measurement. Massachusetts: Allyn and Bacon, Inc.
Aksara.
Prentice-Hall, Inc.


