

Artikel The Analysis of Seventh Grade

by Gusti Nur Hafifah

Submission date: 11-Nov-2021 10:02AM (UTC+0700)

Submission ID: 1699397397

File name: THE_ANALYSIS_OF_SEVENTH_GRADE_Gusti_Nur_Hafifah_full_paper.pdf (170.81K)

Word count: 4327

Character count: 21534

THE ANALYSIS OF SEVENTH GRADE ENGLISH FINAL TEST INSTRUMENT APPLIED IN SOUTH DISTRICT GRESIK SCHOOLS

Sari, Diana Remita & Hafifah, Gusti Nur

English Education Department Faculty of Teacher Training and Education

Muhammadiyah University of Surabaya

Email: gustihafifah@gmail.com

ABSTRACT: This study focused on the seventh grade English final test based on national curriculum in South district Gresik. This research analysed the test appropriateness regarding to basic competence and students levels, difficulty levels, discrimination power, validity, and reliability of the test item. The object was the students' score at the English final test from three different samples of the south district Gresik population. This study used exploratory sequential design of mixed method. The result showed that the content of English final test content was suitable with the basic competence and assessment of the National syllabus of English subject. But it doesn't meet the validity and reliability aspects. However, each item of the test does not have equal distribution and difficulty level. Only one item of the question has appropriate discrimination power. In conclusion, the test item does not meet complete criteria of good test and need revision in order to have a representative result of students' final assessment. This research suggests that Teachers community (MGMP) who in-charged in creating the final test instrument to have further training on creating valid and reliable assessment, which not only rely on the content validity but also reliability of test items.

Key words: Difficulty level, discrimination power, English final test, test appropriateness, reliability, and validity

INTRODUCTION

Education is the process of self-elaboration, self-power to gets the good person through the efforts of the teachers. One of the important parts in the education competent from teachers is evaluating and it is one of indicator in teaching and learning process. It is similarly with Brown (2004:5) that assessment, test and measure are the parts of teaching process. Students' assessment is the report from teaching and learning process to evaluating the method, materials and techniques of education from teachers used. The evaluating teaching and learning process to gets the some feedback and reason to more rises up students' ability and knowledge. Then, evaluating activities of measuring students to know the extent which aims from students achieved of the instructional or students' gets in the learning process (Sudjana, 2011:2). It is also important that assessments do more that the document what students are capable of and what they are knows. Then, to be helpful and useful, assessments should offer evidences as to why students think the way they do and how they are leaning as well as the reasons for misunderstandings (Darling-Hammond, 2013:3).

The analyze of a midterm English test of the seventh grade-accelerated class at SMPN 3 Malang by Arin Eka Primadani, Gunadi Harry Sulisty (2013). This study was conduct to evaluate the quality of the teacher-made test of the midterm English test of the 7th grade accelerated class at SMPN 3 Malang. It was used the descriptive evaluative of the qualitative and quantitative method to analyzed the data. The aspects was analyzes of this study which of the validity, reliability test, item discrimination, and the effectiveness of the distracters. The finding of this analyzes was the items test is valid 100% and the items is reliable. Even though, the items of the test pack have valid and reliable but the test packs were low aspects.

The second previous study was about designing assessment tools based on the principle of language assessment by Yoneda (2012). This research was conduct in a university Business English II class. The result of this research explained that the design of the assessment did not have the criteria of the good designing assessment based on the principle of language assessment. The five principles were practicality, reliability, validity, washback, and authenticity. Because the assessment that teacher create for their students in themes of vocabulary quizzes, paper-and-pencil test, and presentation project is not synonymous with testing, it should be continuous throughout the teaching and learning process in the classroom. Teachers can assess students in many ways other than administering test. It is important to apply the five principles to designing assessments tool.

All the explanation above, the reason why the researcher chosen the title for this study is because the researcher want to know the quality of the English final test items that applied in South District Gresik in 2016/2017 academic years. Some the previous research examine the English final test use to analysis the items with one test-pack or compare one schools and others. And the other

hand, the researcher want to qualification the items test pack that use in 43-50 schools (MTs) in South District Gresik with the random sampling data and the six aspects that must have in the good test-pack. The six aspects are the test based on the curriculum and good assessments checklist, difficulty levels, discrimination power, validity, and reliability test. The all aspects are important to know the quality of the test-pack. Based on it, the researcher want to analyze of seventh grade English final test based on national curriculum in South District Gresik to detect the quality of the items test appropriate with the characteristics of a good test.

METHOD

The design used in this research was exploratory squal design of the mixed method because this study to analyze and evaluated the quality of the test-pack of English final test in seventh grade students of the 1st semester academic year 2016/2017 in South District Gresik by use of the qualitative and quantitative data with the test instruments and result of the object in this research. The qualitative analysis was used to evaluate the content of appropriate toward basic competence and students level (ESL) based on the syllabus and criteria of good assessment in ESL based on Brown and Heaton theories. The quantitative analyses were to find the difficulty level, discrimination power, validity and reliability of the test-pack.

The data was get from the schools documents in MTs South District Gresik from the random sampling technique. The samples was included the Mts Irsyadul Adfal in Benjeng, MTs Sunan Ampel in Kedamean, and Mts Tarbiatul Aulad in Cerme. The data resources for this study were the test instrument and the result of the test. It was included the test and blueprint made by the teachers work group and schools work group to develop the test and the students' score. The test consists of fifty questions in the form of multiple-choice format. The materials presented in the test in form of greeting, announcement, dialog of introduction, animals, time, letter, and family. According to Sudjana (2011:134) that the analysis of the items test was included the students' score to evaluate the difficulty levels, discrimination power, validity, and reliability of the test-pack.

The result of students' paper sheets and students' score was analysis used the Ms. Excel application which included the formula to analyzing the difficulty levels and discrimination power based on the students' upper and lower group in three random sampling. It was usage the formula from Sudjana (2011:135-144). And to find out the measure the result of the data, the researcher used the index of difficulty level and discrimination power for multiple-choice format with 4 options adapted from Ross & Stanley table (Sudjana, 2011:135-145). The validity and reliability test based on the students' paper sheets was analyzing using SPSS version 16.0 and the formula from Priyono (2008:12). Then, to analysis the valid and reliable items looks the R-table from Junaidi, to know the coefficient from the items test.

RESULT AND DISCUSSION

The English final test that the researcher analysis was apply in south district Gresik in 2016/2017 academic years. This test consisted with fifty questions of the multiple-choice form. The test was consisting of the five skills, which fulfill of the requirements in basic assessment in ESL. Those were speaking, listening, writing, reading and structure. In the test were created with using the dialog, description texts, fill in the blank, picture, tree diagram to explain the themes which included in the basic competence on the syllabus usage at seventh grade students. The English final test used in 43 – 50 schools (MTS) in south district Gresik. The obtained data the researcher was take from random sampling which requirements schools of each sub-district. The three samples were namely AS for the first sample, BS of second sample, and CS to third sample. The analysis of the data to gets the aim from this study. Moreover, the aspects to analyzing were qualitative and quantitative of the inquiry.

The Appropriateness Regarding to Basic Competence

After get the paper questions test, the writer adapted the form of curriculum checklist from the syllabus and the form of checklist for selection the content and regarding in ESL from Cunningsworth (2008:58-61). The first result was curriculum checklist consists of two standard competences and eight basic competences. Those are points 3 and 4 of the standard competence and points 3.1 until 4.4 of the basic competence. The result of the fifty questions can be distinguished by the competence standard and basic competence from the syllabus in K13 that implementing in 2016/2017 academic years in South District Gresik.

The numbers that included in 3.1 basic competences were numbers 1, 2, 3 with the percentage of 6%. The next of the point 3.2 basic competence, it was explain the identifying the

social function, text structure, and linguistics of the oral and written communication to asking and giving self-information with the requirements in language which used in family of the basic linguistics. The items numbers that fulfill the requirement in the basic competence 3.2 are numbers 7 and 47 of the percentage 4%. The basic competence point 3.3 was categorizing the social function, text structure and linguistics about namely of day, month, time in numeral, date, year according to the content in the real life or the area of school, home and surrounds. The numbers of this basic competence were 4, 18, 19, and 20 with 8%. Then, the basic competence point 3.4. It was classifying the social function, text structure, and linguistics about transactional in oral or written of sharing and giving information namely of animal, thing, and the building in plural or singular word in the every days. Moreover, eight numbers in point 3.4 of the basic competence has 16%, the numbers of it was numbers 21, 22, 23, 31, 37, 38, 39 and 48.

Subsequently, the point 4.1 of the basic competence was explain to organized the interactional text in oral or written communication that included of the topics greeting, farewell, apologize with the social function, text structure, and linguistics in content. This basic competence has 8%. In addition, numbers were 5, 6, 41, and 50. Then, the basic competence 4.2 has 36% of total numbers 18. Those were numbers 8, 9, 10, 11, 12, 13, 24, 25, 26, 27, 28, 29, 33, 34, 35, 40, 42, and 46. The requirements of this basic competence was organize the interactional text in oral or written to gift and shared the self-information to introducing themselves as the member of the family that included with the text structure, content and linguistics. The point 4.3 of the basic competence description about the organization of interactional text in oral or written communication to sharing and giving information of day, month, time, year, numeral, date with the text structure, linguistics, and content of language. Moreover, for this basic competence, there were ten numbers. Those were numbers 14, 15, 16, 17, 30, 32, 43, 44, 45, and 49. The percentage was 20% for this basic. The last basic competence was point 4.4. It was about the organized the interactional text in the oral or written communication of the information about animal, things, and building in our life. This basic had 2% for one number that includes was number 36.

The Appropriateness Regarding To Students Level (ESL)

The good assessment checklist was to know what the items test appropriateness with the standard in ESL or not. The items test was good for apply in ESL. The researcher analyze from the fifteen number of the good assessments checklist, only five points do not fulfill the requirement. Moreover, 11 points in the items test made by MGMP in South District Gresik. Furthermore, the ten points make the tools of assessment can use to test the students' ability. In addition, the point's numbers 4, 8, 10, and 14 are not necessary. The detailed explanation as follow the table below:

Table 4.3 The Result of Test Appropriateness Toward students Levels

No.	Characteristics of Good Multiple Choice	Checklist (√)
1.	Presented out of simple context	√
2.	The Stem; the choices from which the students select their answers are referred to as option/responses/alternative	√
3.	The option in multiple choice are: a. One option is correct option or key b. Other option is distractors	√
4.	Multiple choice items are all receptive, or selective, responses	X
5.	Design each item to measure a specify objective	√
6.	State both stem and options as simply and directly as possible	√
7.	Make certain that intended answer is clearly the only correct one	√
8.	Use item indices to accept, discard, or revise items	x
9.	Not need long time to answered the question or choose the correct one.	√
10.	All the multiple choice items should be at a level appropriate to proficiency level of education	x
11.	The questions or statements should be grammatically correct	√
12.	The stems should not give clues or question to other question.	√
13.	Chronological on the optional answer	√
14.	Only one feature at the time should be tasted	x
15.	Pictures, graphics, tables, and diagrams should be clear and in function	√

Based on the appendix 4.3 above, the researcher had finished the analyzing of the checklist of a good assessment in ESL. Indeed, the items test was present of the language and the themes were necessary with the matter and the levels of test-taken. The stem, the context of the question has the option / responses / alternative of choices from which the students select their answer and it has on this test-pack. The test-pack has the one correct answer and the others options of the distractor as the false answer. Then, the design for each item was to measure students' ability. Statement both the question (stem) and the options have the connection of the structure text. In addition, it has simply and clearly sentence for the students understand the meaning of the test. It not difficult for students chooses one correct option. According to the test-pack, the writer was thinking that not need must time to students answer the each questions. The grammar of each question is correctly and it has the structure content. Then, the each question not makes students thinking other options except in the option in the paper test. The picture and the chronology each questions was detail and clearly to students understand.

The Difficulty Levels

The Items that can be classified into the difficult category that were six numbers of questions, namely the matter of numbers were 8 (0.036), 19 (0.11), 24 (0), 26 (0.036), 28 (0) and 47 (0.071). The next was the items that had easy as category. The numbers of questions that can be categorized were 44 numbers, and it was number 1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 21, 23, 25, 27, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 48, 49 and 50. The difficulty level item from this sample was **0.77** that means the items test-pack from the category in this sample was easy. Then, the most index of hard level is **0**, and the index of easy level is **0.96**. Therefore, why the items test-pack this sample as call as easy categories.

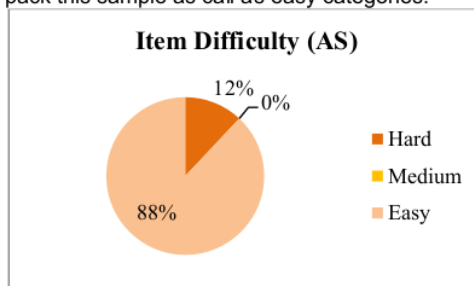


Figure 1 - Figure 4.4 the Result of Difficulty Level in the A Sample (AS)

On the figure above, the writer sum up that items test included the requirement good level of the test did not visible in this data. It where the data analysis of difficulty levels has indicated two categorize from three categorizes must have in the items test. The two categories were easy and hard categories. The percentages of each categorizes were 88% for easy and 12% for hard levels. According to it, the data has not necessary. The levels did not balance in the items test pack. The items test pack must have three categorizes that included in test items, those were medium, easy and hard levels with the ratio 3:4:3 and it means that the percentages of all level as like as 30% for easy, 40% for medium, and 30% for hard level.

From the outcome of the writer get of the B sample, the items difficulty level can be classified like as the most index of hardly category in 0 (Zero), the most medium category in 0,37 (zero point thirty seven), and the last was the most index of easily category in 1 (one). Thus, by the finding data as of BS sample it can be concluded that the mean value for the BS data in 0.67 and it as call as the medium category.

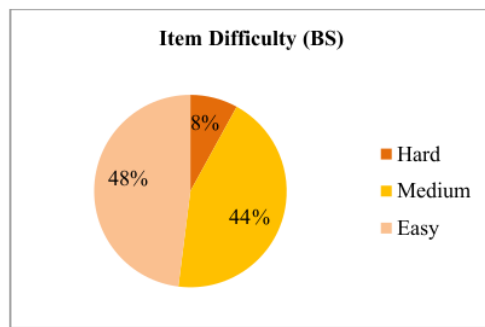


Figure 2 - Figure 4.4 The Result of Difficulty Levels in the B Sample (BS)

The second figure was description of the B sample (BS) in the items difficulty level, which have three categories in the items test. The hard categorize has 8%, medium in 44%, and easy categorize with 48%. The items test did not harmonies because the ratio of each categorize have not appropriate with the requirement of the ideal a good test pack. To sum up, the writer get which hard level was less than others levels. This analyzes from the data students' answer of fifty questions for 30 students. The diagram above concluded that the medium and easy levels were dominant in this analysis with the index of hard level in zero (0), medium in 0.37, and the last level was easy with one (1) of the index.

The data analyzes from the table above, the writer has the percentage of this third sample (CS) with the three levels of difficulty from the 37 students as the test-taken. It was as follows:

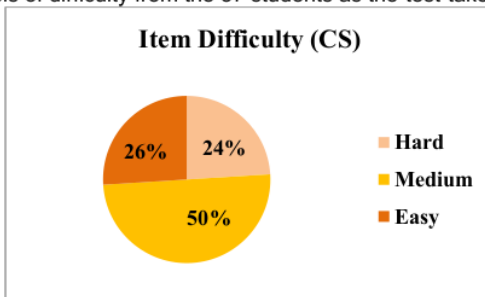


Figure 3 - Figure 4.4.3 The Result of Difficulty Levels in the C Sample (CS)

Based on the diagram above, the researcher can conclude that the items test pack from the sample CS in medium categorize with the percentage of 50%. The easy categorize more than hard categorize on the percentage of 26% > 24% from the fifty questions. Indeed, this items test based on this data is equal with the three categorize that must be had of the medium, easy and hard categorized. The writer get the results which the items test have not the appropriateness of the qualification in the items difficulty levels, because the items tests do not the balance in the categories of the ration 3:4:3 that meaning the hard and easy categories same and the hard categorize was the high value more than others.

The Discrimination Power

The discrimination power of the items test based on the random data, there was one number has on the each sampling. It was number 39, but the indicator of the rank in the upper group each sampling was different. In the AS only has number 39 with the value 5, the BS has nine numbers with different indicators for each numbers, and in the number 39 the indicator is 6. The CS has the higher total numbers of the discrimination power with 13 numbers, and one of the item number was 39 in the indicator 5. It is the indicator same as in AS. Indeed, the item test pack has the discrimination power. The instrument of the test pack was enough to say it. The test packs can usage repeatedly in the next test in south district Gresik.

The Content of Validity

The writer had finished analyzing of the three samples of south districts Gresik in the English final test items in academic years 2016/2017 with the items test pack was fifty questions. Then, the significant of the alpha in validity test was r calculation in two sampling more than r table. Therefore, one sample did not identify with invalid items. The sample is the first sample or AS school. Nevertheless, the total numbers of items is not balance or more low than the interpreted the researcher, but the items test-pack can be said was valid. The all the data were AS sample has one valid items, BS had 16 numbers of the valid items, and CS sample had 12 numbers of the valid items. Indeed, the entire items test was valid with the result analyzes in three sample.

The Reliability

The writer had the result of analyzing that the three samples must have the reliability test. Form the two test-taken said reliable. Those were BS and CS sample. Then, one sample of the test-pack did not reliable. It was the first sample or AS sample. The AS have the r calculation less than r table ($0.129 < 0.3739$). In others wise, the BS has r calculation more than r table with the alpha $0.752 > 0.3610$ and the CS is reliable too. The CS has 0.660 r calculation more than 0.3246 of the r table. Indeed, the all of the items test was reliable, because there were two samples that had the requirements of the reliable items test. And it test-pack can usage in represent in the next test-pack with replacement the sentences of the items test with the same themes or matter in south district Gresik.

CONCLUSION

Based on the analysis of English final test for seventh grade students in South District Gresik at 2016/2017 of academic year, the test was consider have low quality in the good test. In contrast, there was logical relationship between the results of difficulty level, validity and the students' score. The result of difficulty levels in the test-pack has the low quality, divided the category each levels did not equally with fifty questions. There were 4 number questions which fulfill the requirement in the hard level with 8%, 22 number questions for the easy level with 44%, and the medium level with 48% from 24 number questions. The difficulty level was low but the students' score was also low. Then, the discrimination power in the test-pack did not show for all items test, but in one number question. Number question of 39 has the requirement of the discrimination power according to the outcome of the analysis from the samples.

Subsequently, the test which has good quality in measuring the students' achievement, the test should had the appropriateness regarding to basic competence and students level (ESL) to have good validity, reliability, difficulty levels, and discrimination power. Moreover, the test-pack did not have the good assessment, because the items test did not have the criteria and the quality. And the aspects of difficulty from these items have not equally, because according to Sudjana's theory that the hard and easy levels need to balance with the ratio 3:4:3 that means the hard and easy level had the ration 3 with 30% from the fifty questions, and the medium level with 40% from the fifty questions. The discrimination power only in 39 of number question from the three random sampling. From the two aspects the test can say the good designed and instrument to measure the students' ability. The last aspects were validity and reliability. The test-pack had 29 numbers questions of the valid items and 100% reliable. From the all the results, the researcher can concluded that the students in South District Gresik in 2016/2017 of academic year based on the result showed that the difficulty level was low but the students' score was also low. It was prove that students' ability could not complete those questions. Even though having validity and reliability data, only one question which had discrimination power. The items need to modify and improve to be a better test. The teachers must know and understand of the important aspects to designing and created a good test items. Because, the test-pack should be measure the students' ability.

REFERENCES

- Brown, H. Douglas. 2004. *Language Assessment: Principles and Classroom Practices*. San Francisco: Logman Inc.
- Darling-Hammond, et al. 2013. *Criteria for High-Quality Assessment*. June. Chicago: Stanford Center for Opportunity Policy in Education, Stanford University.
- Departemen Pendidikan Nasional Republik Indonesia. 2008. *Standar Pengembangan Kelompok Kerja Guru (KKG) Musyawarah Guru Mata Pelajaran (MGMP)*.
- Heaton, J. Brian. 1990. *Writing English Language Tests*. New York: Longman Inc.
- Heywood, John. 2005. *Engineering Education Research and Development in Curriculum and Instruction*. The United States of America: IEEE Press

- Huges, Arthur. 2003. *Testing for Language Teachers*. UK: Cambridge University Press.
- Junaidi. 2010. Table r (Koefisien Korelasi Sederhana) $df = 1 - 200$.
<http://junaidichaniago.wordpress.com>
- KEMENDIKBUD. 2016. *Silabus Mata Pelajaran Sekolah Menengah Pertama / Madrasah Tsanawiyah (SMP/MTs) Mata Pelajaran Bahasa Inggris*. Jakarta
- Primadani, Arin Eka and Gunadi H. Sulisty. 2013. *An Analysis of A Midterm English Test of The Seventh Grade Accelerated Class At SMPN 3 Malang*. English Language Education. Vol. 3/1 (p: 1-11).
- Priyanto, D. (2008). *Mandiri Belajar SPSS (Statistical Product and Service Solution) untuk Analisa Data dan Uji Statistik*. Yogyakarta: Mediakom
- Sudjana, DR. Nana. 2011. *Penilaian Hasil Proses Belajar Mengajar*. Bandung: Remaja Rosdakarya, Pt.
- Yoneda, Mitaka. 2012. *Designing Assessment Tools: the Principles of Language Assessment*. Japan: Department of English Communication. Vol. 60 (p: 41-49).

Artikel The Analysis of Seventh Grade

ORIGINALITY REPORT

9%

SIMILARITY INDEX

9%

INTERNET SOURCES

0%

PUBLICATIONS

2%

STUDENT PAPERS

PRIMARY SOURCES

1

repository.um-surabaya.ac.id

Internet Source

4%

2

eprints.iain-surakarta.ac.id

Internet Source

2%

3

libir.mukogawa-u.ac.jp

Internet Source

1%

4

Submitted to University of Melbourne

Student Paper

1%

5

jurnal-online.um.ac.id

Internet Source

1%

Exclude quotes On

Exclude matches < 20 words

Exclude bibliography On