

ITEM ANALYSIS OF ENGLISH TEACHER-MADE TEST

Irmala Sukendra

English Education Department, Syekh Yusuf Islamic University Tangerang
isukendra@unis.ac.id

Abstract

Evaluation is a systematic process of determining the extent to which instructional objectives are achieved by the students. Carefully collected evaluation data helped teachers understand the learners, plan learning experience and determine the extent of which the instructional objectives are achieved. This research is aimed to analyze the English summative test validity used by SMA Yaspita, Serpong Utara, Tangerang. The study focused on finding the appropriateness of the summative test, whether it is in line with the syllabus and learning objectives. Thus, this study aimed to find out: (1) the appropriateness of the English material coverage to the content standard of the latest syllabus and (2) the quality of the English summative test items. This study included the English teacher and the students of SMA Yaspita as the considerations of the findings. This research is a descriptive analysis method by using English summative test. The findings of the research discovered that the quality of English summative test at SMA Yaspita of second grade students failed to fulfill the desired learning objectives.

Keywords: content validity, construct validity, teacher-made test

Introduction

Evaluation is needed in every teaching and learning activity to measure students' understanding on the materials given and to ensure that teaching objectives are achieved. An evaluation helps teachers to find out whether or not the learning is succeeded. Information from evaluation has been very useful for teachers to measure student progress or their overall teaching performance in order to achieve the intake of the given materials. In conducting a test, a teacher must follow systematic procedures such as planning the tests, constructing the test items accordingly, trying the test items to ensure the test reliability and validity, administering the tests, assessing the tests objectively, and evaluating the test quality (Gultom, 2017). Evaluation cannot be done spontaneously; it must be structured systematically in order for the learning goals to be achieved. Hence,

ideally, a test should assess what the teaching learning has set its objectives as for the benefit of the students.

In Indonesia, education is centrally controlled by the National Ministry of Education. 6 years of elementary school and 6 years of high school. Elementary levels follow optional pre-school playgroups, which can begin at the age of approximately three. Public schools are administered by the local government and follow the Indonesian curriculum. Conversely, private schools, which are privately funded, generally offer a curriculum that both meets and exceeds the requirements of the Indonesian curriculum. A lack of legal framework for what private and international schools are allowed to set as benchmarks can sometimes lead to a disparity in the quality of the education. The positive value of private schools is that a larger

teacher-student ratio reduces class size and makes it easier for teachers to teach. Private school policies are also more flexible, so necessary courses can be easily added. Parents with bigger concern for their children's education, may put this aspect into consideration when enrolling their children to schools.

However, some private schools may have their own problems as not all teachers are familiar with the implementation of the curriculum and some students are more concerned in getting their degrees and not the education. In term of the case of SMK Yaspita Tangerang, the students appeared to be less concerned with the exam questions they had for English subject. To ensure this issue the writer conducted a structured interview with 10 alumni from the school and discovered that the students were not benefited by the English lesson they got during school years and were not assessed properly based on their competence.

Kathryn et al. (2017) stated that teachers should conduct a good and proper test in order to accurately measure students' comprehension. Yet, based on the writer's observations at SMA Yaspita South Tangerang, some test questions were taken from random sources on the internet. Although the English teacher, Mrs. K, claimed that there were no problems with the students regarding the exam questions that did not correspond to the knowledge that had been taught (personal communication, Dec 20, 2022). Teachers play a significant role in creating assessments and tests for their students. Due to this, teacher-made tests are highly regarded, as they are designed to meet the unique needs of the class and align with the

curriculum. However, teacher-made tests can often lack reliability and validity, which may affect students' comprehension of the lesson. In relation to that, Al-Hudawi et al. (2014) indicated that "major absence across the language teaching was found in 'developing the potentials of individual in accordance to student perspectives'" (p. 65). In line with that, Puspendik Kemendikbud issued in 2019 of the Indonesian Centre for Educational Evaluation (as quoted in Yanti et al., 2020, p.36) suggested that if the items of the test do not meet the specifications, the test is considered to be of poor quality. Therefore, any tests, especially one manufactured and implemented in Indonesia, should follow the practice of manufacturing good test to maintain quality. Thus, this study aims to analyse the appropriateness of the English material coverage to the content standard of the latest syllabus and the quality of the English summative test items in SMA Yaspita, South Tangerang.

Teacher-made test

Teacher-made tests are designed by the teacher based on the syllabus and lesson plan applied in the course. It is designed to measure the success rate of students in meeting course objectives in the teaching process carried out by teachers. The teacher cares more about the test created by the teacher because he is directly involved in the creation of the test. A teacher-made test is constructed to measure outcomes directly related to classroom specific objectives and particular class situations. Therefore, teachers must make logical and

rational questions about what items are worth asking. This test is usually used for daily, formative, and general test (summative) tests (Brown, 2018, p. 66). The effectiveness of such tests depends on the teacher's experience and ability to create the tests. This test is within the ability of each teacher and is also the most economical.

Many teachers can provide students with good teaching materials and test, but most of them do not analyse the quality of their own tests, including confidence coefficient, difficulty index, strength index and the distractor index of the tests. They are still lack of awareness about the importance of test analysis. Time constraint and other responsibilities force the teachers not to analyse their test items referring to several characteristics of good test item (Lebagi, Sumardi, & Sudjoko, 2017, p. 98). Brown (2018, p. 115) mentioned that basically the teacher-made test can only be used in some classes the teacher teaches. The advantage of using this kind of test is that students are familiar with the task given by the teacher enabling them to get a better score than in standardized test. To name a few, studies from Kheyami et al., (2017), Paramartha (2017), Sahoo and Singh (2017), Rehman, et al., (2018), Mahjabeen et al., (2018), Kusumawati & Hadi (2018), and Rao, et al., (2018) addressed the teacher-made test's quality from the item discrimination, item facility and distractor efficiency

The limitation of teacher made tests are limited sampling, low reliability, subjective, low validity, high skill required, monotonous, spend lots of time to compose. A test

should reflect the teaching, test becomes a tool to assess what students have been taught. A test should ultimately question what the students have been taught. Test is aimed to assist teacher to determine the success of teaching and aid them to improve their teaching quality. According to Arikunto (2021, p. 29) a teacher-made-test is a test that is produced and created by a teacher in the classroom, hence the test's validity and reliability are not the same as standardized tests. Because effective teachers review and amend test items that have been tested, they are unaware of their level of validity and reliability. This test is based on materials and specific objectives devised by the teacher for his own students.

Method

In this research, there are several data collection techniques used, including data collection documentary with regard to the summative test questions sheet and interviews with users. The writers used eight (8) approaches to examine the validity of this English summative test based on Fahriany et al. (2018) which are: (1) identifying, that is locating relevant facts required for records evaluation, (2) deciding on or determining which parts need to be analysed, (three) coding or identifying each kind of facts, (4) creating a tick list, that is categorizing the data based at the decided traits, (5) tabulating or entering data primarily based on type, (6) analysing, providing an intensive description of the studies trouble, (7) interpretation that's

developing study findings which are then compared to different similar studies, and (8) conclusion or summarizing the research findings, making inferences, and responding to the studies questions. For this observation, the writers used the English test for the eleventh-grade students.

The writers used descriptive analysis of the test items for each range in reading the information to identify the specific facts about the test that conformed to the syllabus, test indicators and learning objectives. The test has 40 items with 4 options provided in each number. The writers began by analysing all of the test items. the writer then categorized the questions based on the base competence of 2013 curriculum. The next step was to explain the conformity of each test item to the indicators of each base competence.

The writers used the distribution frequency relative method to decide the percentage of items that adhered to the syllabus. This step is done to obtain the appropriateness of the test. To achieve the relevancy of the test, the writers interviewed the test takers (the students) and asked them about the test items relevance. This step was also done by cross checking the test items with the learning objectives proposed by the teacher.

Tabulation of conformity of each test item was done once the list of quality was achieved in order to ensure that the test was properly administered.

Findings and Discussion

There are 12 sub themes of teaching material that are being taught at SMA Yaspita based on the syllabus which cover: suggestions and offers, opinions, hopes and prayers, formal invitation, personal letter, procedure text, passive voice, conditional sentence, factual report text, analytical exposition text, biography, and song. Based on those themes, two (2) of them which are hopes and prayers; and songs are not assessed in the summative test. Meanwhile, the other 10 themes are considered as the learning contents, are assessed in the test.

The summative teacher-made test has 40 test items in the form of multiple choice. The writers discovered that there were 22 test items that are in line with the objectives of the syllabus and 18 which are not in line with the syllabus. The following table describes the test distribution which conformed the syllabus requirements:

Table 1. Summative Test Item

Subject matter	Items number	Total
----------------	--------------	-------

Spoken and written texts to make suggestions and offers and their responses. (examples of caring, cooperative, and proactive behavior) Expression Suggestions and offers: <ul style="list-style-type: none"> • Why don't you • What about ... ? • You should... • You can... • Do you need ... ? 	3	1 Item
Oral and written texts to express opinions and thoughts and their responses Expression Express your opinion/thoughts <ul style="list-style-type: none"> • I think.... • I suppose.... • In my opinion 	17	1 Item
Oral and written texts to express hopes and prayers and their responses Expression Hope and pray <ul style="list-style-type: none"> • I hope ... • I wish you all the best. Thank you 		
Special text, spoken and written, in the form of a simple formal invitation Structure Salutation <ul style="list-style-type: none"> • will/could you come with me to the exhibition? • Is it possible for you to attend my birthday party ? 	37,38	2 Items
Simple, personal letter Structure <ul style="list-style-type: none"> • Date • Salutation : Dear ... • Opening Paragraph : • Greetings and share the current situation and what is being done • Content: report things that have happened/has been happen • Closing: closing the letter in the hope of meeting again Signature	10,11, 18, 19,20	5 Items
Procedure text in the form of manuals and tips (tips) structure mention the material/part of the object that is described in full, as well as a list of steps to be taken	36	1 Item
Actions/activities/events without mentioning the culprit (passive voice) Text structure <ul style="list-style-type: none"> • Insects are considered dangerous animals. <ul style="list-style-type: none"> • Tsunami is caused by earthquake affecting the seabed 	21,22,23	3 Items
Supposition if something happens in the future (conditional sentence) Text structure <ul style="list-style-type: none"> • If teenagers eat too much fast food, they can easily become 	5,6	2 Items

<p>overweight.</p> <ul style="list-style-type: none"> If you exercise regularly, you has been get the benefit physically and mentally 		
<p>factual scientific text (factual report) simple oral and written about objects, animals and natural phenomena/events Structure General classification of written animals/objects, e.g. Slow loris is a mammal. It is found in it is a nocturnal animal. It is very small with A description of the part, its nature and behaviour</p>		
<p>Analytical exposition text Text structure State the subject matter of something that is hotly discussed <ul style="list-style-type: none"> State your views/opinions on this matter along with illustrations to support it Ends with a conclusion that restates the opinion on this matter </p>	7,8, 28,29,30	5 Items
<p>About famous figures Structure <ul style="list-style-type: none"> Mention actions/events in general Mention the sequence of actions/events in chronological order, and in chronological order If necessary, there are general conclusions </p>	39,40	2 Items
<p>Song Topic : exemplary behaviour that inspires</p>		
Total		22 Items

$$P = \frac{F}{N} \times 100\%$$

$$P = \frac{22}{40} \times 100\%$$

$$P = 55\%$$

$$P = \frac{F}{N} \times 100\%$$

$$P = \text{Percentage}$$

$$F = \text{Frequency of unconformity}$$

$$N = \text{Number of Sample}$$

Hence, the calculation of the test items that conform with the syllabus is 55% of the total test item questions, whereas the other 45%

unconfirmed with the syllabus (the distribution of test description can be seen in the table as follow:

Table 2. Unconformity between The Summative Test Items and English Syllabus

No	Indicators that are not found in English Syllabus	Items Number	Total
1.	Warning	1,4	2 Items
2.	Expression of surprise	2	1 Item

3.	Expression of relief	12	1 Items
4.	expression of pain	14	1 Item
5.	expression of pleasure	15,	1 Item
6.	vocabulary	9, 27, 34	1 item
7.	conjunction	16	1 Item
8.	narrative text	24,25,26,30,31,32,33,35	10 Items
Total			18

The table shows that almost half of the test items are inappropriate to be tested as they are not required in the syllabus and therefore it should not be taught and tested to students.

The excerpts below show that the test items are not only inappropriate to be tested in terms of the irrelevancy with the syllabus and lesson plans but the test items are also unqualified. Question number 3 requires the students to give an advice since the options could not be

addressed as distracters as there is only one suggestion. However, in communication, options A, C, D, and E are plausible. A speaker may naturally respond “up to you” when they were asked to give their opinion. Similarly, option C and E are also commonly occurred in conversation. The quality of this item is worsened by the grammatical errors that was displayed in both the question and in the options.

No	The Question
3	<p>Give advice for this problem with the correct expression of offering advice! problem. I don't have money to pay my English book. I'm afraid to asking it th my father because he sick and has not job. What should I do?</p> <p>a. Up to you b. I suppose so c. That is not my affair d. I suggest you to get job e. I don't have opinion</p>

Grammatically the test is better written as:

Give advice for this problem with the correct expression!

I have a problem. I don't have money to pay FOR my English book. I'm afraid to asking FOR it ~~with~~ FROM my father because he IS sick and has not GOT ANY job. What should I do?

- a. Up to you
- b. I suppose so

- c. That is not my ~~affair~~ PROBLEM
- d. I suggest you ~~to~~ get job
- e. I don't have ANY opinion

Question number 6 asks about the action the students would text under certain condition, which means pragmatically, the students are free to express their opinion and therefor any answer they gave is correct. The question also has ungrammatical sentence for its options. Not to

mention that option B and D could be interpreted similarly.

6.	What would you do in this situation? You're 20 minutes late for class The teacher is explaining something to the class when you arrive What would you do?
a.	Not go to the class
b.	Say nothing and take a seat
c.	Go in, walk up to the teacher and apologize
d.	Go in as quietly as you can and take a seat
e.	Wait outside the classroom until the class

Question number 17 has clearly wrong answers for option A and C as one cannot agree with a thing and the expression "do you know 'X'?" is used to introduce someone. Option D

is confusing as the sentence can be corrected in two ways: (a) Have you complained about this refrigerator? Or (b) Have you any complaint about this refrigerator?

17.	Complete the following dialog with suitable expression of asking opinion!
a.	Do you agree with this refrigerator?
b.	What do you think of this refrigerator?
c.	Do you know this refrigerator?
d.	Have you complain about this .refrigerator?
e.	Are you satisfied with thisrefrigerator?

Question number 28 requires the students to demonstrate their vocabulary and grammar knowledge whereas the test item is

grammatically wrong. 'Has been' needs to be followed by Verb + ing, and there is not any of the options that has V+ing construction.

28.	If you withdraw your money from the cash dispenser the amount of your money has been ... as you have drawn out.
a.	Become more
b.	Be credited
c.	Become lost
d.	Become less
e.	Be doubled

Conclusion

Based on result of the items analysis it can be concluded that the English summative test which administered

in SMA Yaspita cannot be claimed as a good tool to assess the students' competence and knowledge. The test failed to assess the competencies that were outlined on the syllabus and the lesson plans for the subjects. Some materials that are tested are not included in the syllabus. The teacher did not adjust the test questions with the materials that had been taught and to the government curriculum through the lesson plan. Those are question on warning in questions number 1 and 4, expression of surprise in question number 2, expression of relief in questions number 12 and 13, expression of pain in question number 14, expression of pleasure in question number 15, conjunction in questions number 16, and narrative text in questions number 24 through 27, and 30 to 35. There are also many grammatical mistakes in test that it raised another question of the teaching quality. Thus, it can be inferred that the quality of English summative test at SMA Yaspita of second grade students failed to fulfill the desired learning objectives.

This study is a glimpse reflection of Indonesian education. There could be other cases where teaching is simply a fulfillment of going to classes without getting the education. The quality of Indonesian students is in questions. I do believe that there are many 'good' and qualified schools, but the real question is: are they available for everyone? Are they accessible for all or exclusively existed for some people? How does the government

improve the quality of Indonesian education? Perhaps the problem is not only about the regulations and policies.

References

- Aaron S. Richmond. (2016). Constructing a Learner-Centered Syllabus: One Professor's Journey. *Idea*, 60(September).
- Adom, D., Mensah, J. A., & Dake, D. A. (2020). Test, measurement, and evaluation: Understanding and use of the concepts in education. *International Journal of Evaluation and Research in Education*, 9(1), 109–119. <https://doi.org/10.11591/ijere.v9i1.20457>
- Ahmed, J. U. (2010). Documentary Research Method: New Dimensions Jashim Uddin Ahmed 1. *Indus Journal of Management & Social Sciences*, 4(1), 1–14.
- Ajayi, O. V. (2017). Candidate Names: Distinguish between Primary Sources of Data and Secondary Sources of Data, 1–5.
- Aldurayheem, A. (2022). The effectiveness of the general aptitude test in Saudi Arabia in predicting performance of English as a foreign language. *PSU Research Review*. <https://doi.org/10.1108/prr-01-2021-0002>
- Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. *Research in Social and Administrative*

- Pharmacy, 15(2), 214–221.
<https://doi.org/10.1016/j.sapharm.2018.03.066>
- Aloud. (2017). Journal of Educational Research and Evaluation. Jurnal of Educational and Evaluation, 6(1), 10–18.
- Bahri, M. F. (2019). Muhammad Fajrul Bahri. ISLAM Realitas: Journal of Islamic & Social Studies, 5(1), 45–51.
- Baldwin, S., & Cheng, L. (2020). Internationally Educated Nurses and the Canadian English Language Benchmark Assessment for Nurses: A Qualitative Test Validation Study of Test-Taker Accounts. Canadian Journal of Applied Linguistics, 23(2), 96–117.
<https://doi.org/10.37213/cjal.2020.30435>
- Designing an English Language Syllabus for Teacher Education Programs in Indonesia. (2017). Elementary School Journal Pgsd Fip Unimed, 7(3), 329–336.
<https://doi.org/10.24114/esjpgsd.v7i3.8168>
- Faridi, A., Bahri, S., & Nurmasitah, S. (2016). The Problems of Applying Student Centered Syllabus of English in Vocational High Schools in Kendal Regency. English Language Teaching, 9(8), 231.
<https://doi.org/10.5539/elt.v9n8p231>
- Furwana, D. (2019). Validity and Reliability of Teacher-Made English Summative Test at Second Grade of Vocational High School 2 Palopo. Language Circle: Journal of Language and Literature, 13(2).
- Gultom, E. (2016). Assessment And Evaluation In Efl Teaching And Learning. Assessment and Evaluation in Efl Teaching and Learning, 190–198.
- Ismail, T. B. M., & Alsahhi, N. R. (2022). Language Assessment and Writer Identity in Higher Education: A Qualitative Analysis of Labeling Learners through Placement Tests. Information Sciences Letters, 11(2), 417–426.
<https://doi.org/10.18576/isl/110212>
- Kassab, S. E., Bidmos, M., Nomikos, M., Daher-Nashif, S., Kane, T., Sarangi, S., & Abu-Hijleh, M. (2020). Construct validity of an instrument for assessment of reflective writing-based portfolios of medical students. Advances in Medical Education and Practice, 11, 397–404.
<https://doi.org/10.2147/AMEP.S256338>
- Kathryn, M., Anita, R., Dewi, R., & Kristiandi. (2017). Rapid Review of Curriculum 2013 and Textbooks. Education Sector Analytical and Capacity Development Partnership (ACDP), 186.
- Kibble, J. D. (2017). Best practices in summative assessment. Advances in Physiology Education, 41(1), 110–119.
<https://doi.org/10.1152/advan.00116.2016>
- Koller, I., Levenson, M. R., & Glück, J. (2017). What Do You Think You Are Measuring? A Mixed-Methods Procedure for Assessing the Content Validity of Test Items and Theory-Based Scaling. 8(February).
<https://doi.org/10.3389/fpsyg.2017.00126>
- Mathieu, J. E., Tannenbaum, S. I., Kukenberger, M. R., Donsbach,

- J. S., & Alliger, G. M. (2015). Team Role Experience and Orientation: A Measure and Tests of Construct Validity. <https://doi.org/10.1177/1059601114562000>
- Menteri Pendidikan Dan Kebudayaan Republik Indonesia. (2013). Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 22 Tahun 2016 tentang Standar Proses Pendidikan Dasar dan Menengah. Kementerian Pendidikan Dan Kebudayaan, 53(9), 1689–1699.
- Mohammed, S. W., & Abdurehman, O. S. (2020). Exploring challenges of preparing authentic English language tests in secondary and preparatory schools: The case of Gamo, Gofa and Konso zones in Southern Ethiopia. *Educational Research and Reviews*, 15(8), 433–437. <https://doi.org/10.5897/err2020.3965>
- Permendikbud. (2018). Permendikbud RI Nomor 37 tahun 2018. *JDIH Kemendikbud*, 2025, 1–527.
- Putri, B. D. T. (2018). the Validity Analysis of English Summative Test of Junior High School. *Journal of Languages and Language Teaching*, 5(1), 6. <https://doi.org/10.33394/jollt.v5i1.327>
- Richmond, A. S., Morgan, R. K., Slattery, J. M., Mitchell, N. G., & Cooper, A. G. (2019). Project Syllabus: An Exploratory Study of Learner-Centered Syllabi. *Teaching of Psychology*, 46(1), 6–15. <https://doi.org/10.1177/0098628318816129>
- Rohmah, N. (2019). Validity and Reliability Study on Teacher-Made Assessment for English Mid-Term Examination. 254(Conaplin 2018), 107–110. <https://doi.org/10.2991/conaplin-18.2019.236>
- Saeed Alharbi, A., & Meccawy, Z. (2020). Introducing Socratic as a Tool for Formative Assessment in Saudi EFL Classrooms. *Arab World English Journal*, 11(3), 372–384. <https://doi.org/10.24093/awej/vol11no3.23>
- Shorten, A., & Smith, J. (2017). Mixed methods research: Expanding the evidence base. *Evidence-Based Nursing*, 20(3), 74–75. <https://doi.org/10.1136/eb-2017-102699>
- Shrotryia, V. K., & Dhanda, U. (2019). Content Validity of Assessment Instrument for Employee Engagement. *SAGE Open*, 9(1). <https://doi.org/10.1177/2158244018821751>
- Siddiek, A. G. (2010). The Impact of Test Content Validity on Language Teaching and Learning. *Asian Social Science*, 6(12), 133–143. <https://doi.org/10.5539/ass.v6n12p133>
- Sugianto, A. (2017). Validity and reliability of english summative test for senior high school. *Indonesian EFL Journal: Journal of ELT, Linguistics, and Literature*, 3(2), 22–38. <http://ejournal.kopertais4.or.id/mataraman/index.php/efi/article/view/3191/2432>
- Sumarni, W., Susilaningsih, E., & Sutopo, Y. (2018). Construct Validity and Reliability of

- Attitudes towards Chemistry of Science Teacher Candidates. *International Journal of Evaluation and Research in Education (IJERE)*, 7(1), 39. <https://doi.org/10.11591/ijere.v7i1.11138>
- Yaghmaie, F. (2003). Archive of SID Content validity and its estimation Archive of SID. *Journal of Medical Education*, 3(1), 25–27.
- Yao, D., Liao, L., & Ye, X. (2022). Investigating the Construct Validity of an ESL Test for Young Learners Investigating the Construct Validity of an ESL Test for Young Learners. May. <https://doi.org/10.5430/wjel.v12n5p59>
- Zaitolakma, N., Samad, A., Husin, N., Zali, M. M., Mohamad, R., Che, A., Malaysia, M., Terengganu, U., & Campus, D. (2018). A Correlation Study on Achievement of English Learners. *Mjltm*, 8(3), 407–430.