

## COMPREHENSIVE TEST ITEMS ANALYSIS OF TEACHER-MADE ENGLISH SUMMATIVE ASSESSMENT IN THIRD GRADE PRIMARY SCHOOL

*Enggar Pangesti Wibowo*  
*Universitas Pendidikan Indonesia*  
*enggarpangestiw@upi.edu*

Submitted: 2024-12-14

Accepted: 2025-03-07

---

**Abstract:** Summative test is a written test given at the middle or end of the semester. This study aims to investigate the overall quality of a teacher-made English summative test at one of private Islamic elementary school in Semarang, Central Java, using the general and specific guidelines set for constructing good test items. This research offers a document analysis of a qualitative study involving teacher-made test at primary school level that remains under-explored. The study reveals that the test instrument consists of single and complex multiple-choice questions, short answers and essays that assess the students' ability in vocabulary, grammar, and simple sentences. The test items generally meet most of the aspects of the principles, but some aspects require further attention such as the use of correct punctuation and spelling, clear, contextualised and functional images, and appropriate grammar. In particular, the options in single-choice and complex multiple-choice questions should be organised homogeneously. For short answer questions, it is necessary to ensure that each question has one correct answer. On the essay questions, no significant problems were found. Revisions are needed to enhance the quality of the test. Accordingly, this study provides valuable insights and improvement recommendations to address aspects that become concerns.

---

**Keywords:** assessing young learners; summative test; teacher-made test; test in primary school

---

### INTRODUCTION

The term assessment is often interchangeable with test, but they are actually different (Torres, 2019). Assessment can be interpreted as a procedure to obtain information from the process and results of learning carried out by teachers and students towards certain learning objectives/criteria/achievements while tests are more of a form of assessment where all tests are assessments, but not all assessments are tests. Brown and Abeywickrama (2019) add that tests are one component of assessment that serves as a tool to assist teachers in assessing students through various administrative tasks.

Assessment has varying roles and purposes depending on what aspects of students are to be measured. According to Purpura (2016), among the many types of assessments, summative assessments are more frequently conducted. Summative assessments or tests, which are usually administered at the end of a learning period such as end-of-semester or end-of-year tests, are used to evaluate student success and teaching effectiveness. In addition, results from summative assessments are often used for administrative purposes, like graduation determination, student ranking, and accreditation of educational programs, making them an important component in the formal education system.

Assessment at different school levels has a distinctive focus according to the educational objectives at each level. At the primary level, for example, language assessment involves language components of grammar and vocabulary in all test papers to ensure that test takers have a strong and thorough understanding of the basic aspects of the language being learnt (De Silva, 2021). Aside from that, Sarigöz and Fişne (2018) found that writing proficiency and vocabulary are key aspects of language that the teachers thoroughly evaluate in lessons with young learners. Thus, EFL teachers often use written exam papers.

According to Rea-Dickins and Rixon (1999), when asked how they assessed their young learners students, most teachers said that they used grammar and vocabulary tests, fill-in-the-blank, and single sentence exercises in written tests. Furthermore, in assessing young learners, teachers should ensure that each assessment format is introduced with clear and simple instructions. They may want to present it using pictures rather than words (Malloy, 2015) as students can utilize pictures or visual symbols to investigate and expand upon their answers (Beach, 1993). In addition, Malloy (2015) recommends that teachers keep assessment tasks short and vary them frequently.

In evaluating the language skills of young learners, the use of standardized written tests is not recommended. Instead, teacher-made tests that are more suitable to the needs and abilities of lower grades students can be a more effective alternative in assessing their language comprehension. However, one of the drawbacks of teacher-made tests is that their quality can vary and may not necessarily meet the standards set (Stiggins, 1988; Wiggins, 1989).

Researches evaluating the quality of teacher-made tests in English subjects have commonly been conducted at the junior and senior high school levels as done by Afriyanti (2015), Erna et al. (2018), and Septi et al. (2020). Besides, as Benigno and de Jong (2016) believe that constructing a test must be in accordance with the learning objectives that have been set, De Silva (2021) found that there was a mismatch between what was covered in the syllabus and what was assessed in teacher-created assessments. Many teacher-made tests do not fulfil the objectives of the syllabus. At the primary school level, similar researches have not been conducted as the analysis of teacher-made English tests is often done quantitatively, for instance conducted by Lebagi et al. (2017), Marsevani (2022), as well as Mina and Wijaya (2014). This research offers novelty by focusing on the qualitative analysis of teacher-made English tests at the primary school level on grade 3, an area of research that remains under-explored.

Accordingly, this study attempts to seek answers of two research questions. First, it investigates the overall quality of teacher-made English summative test for third grade primary school aligned with established general guidelines. Second, it explores in what ways each type

of questions in the test adhere to or deviate from the recommended specific guidelines for constructing good assessment items.

**METHOD**

This research employed a qualitative descriptive method using document analysis approach. Document analysis entails examining and interpreting data to derive meaning, understanding, and develop empirical knowledge (Bowen, 2009). Creswell (2015) states that document analysis is an excellent source of text data in qualitative research as it is reliable for the purposes of such research.

The data analysed were the English exam items created by the teacher. It was analysed together with the syllabus used by the teacher to ensure compatibility between the material taught and the questions tested. The teacher-made test in this study was an even term summative test in grade three at a private Islamic primary school in Semarang, Central Java.

The data analysis method was carried out by reading each test item rigorously and matching the items with a table of good test writing criteria that the researcher developed. These criteria cover three main aspects: material relevance or content validity, to ensure the content is aligned with the learning objectives; construction, to evaluate the clarity of question formulation and the organisation of answer options; and language, to assess grammatical accuracy and readability for test takers. The developed writing criteria are based on guidelines proposed by Cunningham (1998), Kansas Curriculum Center (2001), and Puspendik (2019).

**FINDINGS**

**The Overall Quality of Teacher-Made English Summative Test Align with Established General Guidelines**

There are 30 test items in the English summative test made by the teacher. It consists of 17 single multiple-choice questions, 5 complex multiple-choice questions, 5 short answers and 3 essay questions that has been uniformed for other subject tests as listed on the school test format. The first part of the findings includes a general analysis of each item type, based on the guidelines applicable to both objective and subjective tests (constructed response item) as a whole. The following table (Table 1) presents the results of the item review covering single and complex multiple choice, short answer, and essay forms.

**Table 1. General Test Writing Criteria**

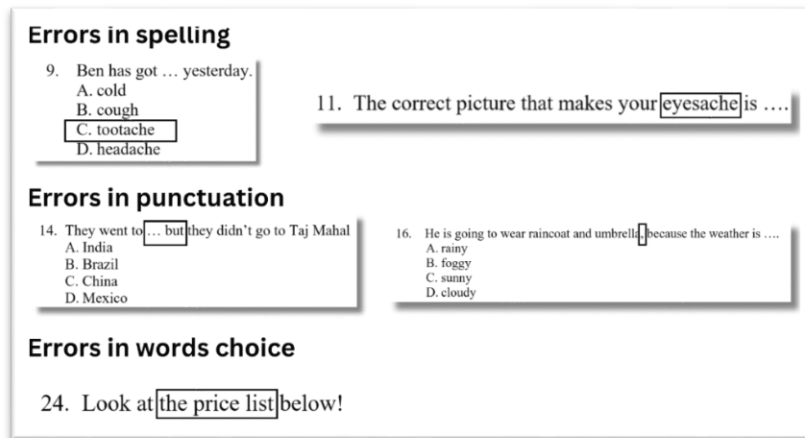
<b>General Aspects reviewed</b>	<b>Total numbers fulfilling the guidelines</b>	<b>Guidelines Fulfilled (%)</b>
1. The instructions in test section are clear and easy to understand.	4/4	100%
2. There are scoring guidelines.	4/4	100%
3. The items match objective content in the curriculum.	30/30	100%

4. The materials asked are in accordance with competence (urgency, relevance, continuity, and daily use).	29/30	97%
5. The content materials are appropriate for the type of school or grade level.	30/30	100%
6. The subject matters are logical in terms of material.	29/30	97%
7. The items clearly formulate a problem.	28/30	93%
8. Test items do not use ambiguity which can be confusing.	27/30	90%
9. Item characteristics don't provide clues to the correct answer.	30/30	100%
10. The items don't provide unfamiliar of the words used.	29/30	97%
11. The items are grammatically correct.	24/30	80%
12. The items have correct spelling, punctuation, capitalization, and word choice.	19/30	63%
13. The correct answers do not depend on the previous options.	30/30	100%
14. Stimuli in the form of pictures, graphs, tables, diagrams, and the like contained in the questions are clear, functional, and contextual.	14/20	70%
15. Test items use communicative language.	29/30	97%

From the table above, it can be seen that the amount of data displayed in the 'Total numbers fulfilling the guidelines' column varies for some of the principles reviewed. This is because some guideline points are only applied to some numbers or subtests as a whole. For example, '.../4' indicates that all four subtests; single multiple-choice, complex multiple-choice, short answer, and essay, fulfil the criteria set. The number '.../20' specifies the test items fulfilled out of a total of 20 questions equipped with pictures while '.../30' points out the review of a total of 30 questions.

The finding shows that not all writing criteria are met. There are some errors that are most frequently encountered in the test. The three main errors are errors in the category of spelling, punctuation and capitalization; the use of stimuli in the form of pictures, graphs, tables, diagrams and the like; and grammar.

Issues in spelling, punctuation and capitalization are the most dominant errors in the teacher-made test. There are only 19 out of 30 questions fulfilling the writing rules, with an achievement rate of 63 per cent. This means that the other 11 questions do not fulfil the rules set. Here are examples of some questions that have writing errors, which are outlined in the following figure.



**Figure. Some examples of errors in spelling, punctuation, and word choice**

Test items 9 and 11 are examples of misspelled questions. Examples of errors in punctuation can be seen in test items number 14 and 16 where according to Butterfield (2015) and Garner (2016), the conjunction *but* in number 14 should be preceded by a comma because it connects two independent clauses and the comma in number 16 is not needed before *because* between the two clauses. The phrase *price list* in test item 24 is considered as an error in word choice as it typically refers to a comprehensive list of prices for a range of items, often found in a catalogue or menu. In the context of the test item 24, where there are only two items following the instruction, a different phrase might be more appropriate to avoid any confusion.

The second most common error in teacher-made test is the use of stimuli such as pictures, graphs, tables, and diagrams that are not clear, functional, and contextualised. There are only 14 out of 20 questions met this criterion, with an achievement rate of 70 per cent. In this test, the stimuli used by the teacher are only pictures. However, the pictures are inconsistent, with some in colour and some not.

Grammatical errors are also found in 6 out of 30 items, indicating 80 per cent achievement. Test item 6, for example, is grammatically incorrect as the verb *used* should use the present tense *use*. The question uses present tense to express common habits, so the verb used should also be in present tense. In test item 24, there should be no final *-s* in the word *million* when we talk about exact numbers followed by quantities such as hundred, thousand, and million.

Some minor errors encountered in teacher-made test also include questions with material which is not in accordance with competencies, unclear material logic, unclear problem formulation, ambiguous language use, the use of words that are unfamiliar to students, and the use of language that is less communicative.

Overall, the test question pattern focuses on vocabulary testing and elementary grammar with simple sentences. In addition, it is also found that repetition of words in the stem and

multiple-choice options occur in several numbers. They are test item 5, 6, 9, 10, 13, 14, 15, 16, and 17.

### **The Quality of Each Item Type on The Teacher-Made English Summative Test Based on the Specific Guidelines**

The first section of the test is single multiple choice (17 items) followed by complex multiple choice (5 items). As the two test types are almost similar in their characteristics, the following table (table 2) shows the results of the item analysis of the specific guidelines that combined single and complex multiple choice for a total of 22 questions.

**Table 2. Specific Test Writing Criteria for single and complex multiple choice**

<b>Specific Aspects reviewed</b>	<b>Total numbers fulfilling the guidelines</b>	<b>Guidelines Fulfilled (%)</b>
1. Answer choices are homogeneous.	17/22	77%
2. Test items must have:		
• only one correct answer (for multiple choice)	17/17	100%
• more than one exact correct answers (for complex multiple choice)	5/5	100%
3. The stems are not filled with unrelated content.	22/22	100%
4. Test items do not use of 'all of these' or 'all the above' as well as 'none of the above' as an alternative option.	22/22	100%
5. The length of the answer choices is relatively the same.	21/22	95%
6. The exact correct answer for each number is valid.	22/22	100%
7. The answer choices in each question do not repeat words or phrases that do not constitute a single understanding.	20/22	91%

The table above (Table 2) shows that the multiple-choice test items written by teachers are almost close to good writing instructions based on specific writing rules. However, there are some points that still need to be improved. There are questions with options which are not homogeneous, repetition of words in the answer choices, and unequal length of options.

The next part of the test in short answer. There are 5 questions. The result of item analysis of the specific guidelines is presented in the table below (Table 3).

**Table 3. Specific Test Writing Criteria for short answer**

<b>Specific Aspects reviewed</b>	<b>Total numbers fulfilling the guidelines</b>	<b>Guidelines Fulfilled (%)</b>
1. Test items have one correct answer.	2/5	60%
2. Fill-in questions use statement sentences.	4/5	80%
3. The subject matter does not contain double negative statements.	5/5	100%

4. The blank is put at the end or at the centre of a statement rather than at the beginning.	4/5	80%
5. The answers required by the questions are short and definite, i.e. in the form of a word, phrase, numbers, symbols, places, or times.	5/5	100%

As with the previous forms, most aspects of the writing criteria are well met in the short answer questions. The two areas that require improvement are ensuring each question has only one correct answer and paying attention to consistency in the use of statement sentences for fill-in questions.

Lastly, essay questions on this exam were found in the form of naming vocabulary and arranging scrambled words to form complete sentences. This part which consists of three questions is all clear and has no issue in the specific aspects. It can be seen in table 4 below.

**Table 4. Specific Test Writing Criteria for essay**

<b>Specific Aspects reviewed</b>	<b>Total numbers fulfilling the guidelines</b>	<b>Guidelines Fulfilled (%)</b>
1. The boundaries of the questions and expected answers are clear.	3/3	100%
2. The sentence formulation of the question or question use an interrogative or command word that demands an elaborate answer.	3/3	100%

## **DISCUSSION**

The summative test in this study was made by the teacher by adjusting the format from the technical guidelines for the implementation of examinations in school and the syllabus set out in the curriculum. Although this test has followed good writing rules, there are several things that must be revised. One of the strengths of this test is the variety of question forms presented. Most schools only use multiple choice or a combination with essays. The use of different forms of tests for young learners helps students prepare for various evaluation formats at higher levels of education. For example, the Assessment of Minimum Competence (AKM) in grade 5 primary school integrates various forms of questions such as single multiple choice, complex multiple choice, matching, true false, fill-in-the blank, and essay.

In general, all forms of tests including single multiple choice, complex multiple choice, short answer, and essay are clearly written and easy to understand. These tests have been customised according to the syllabus so that the material taught by the teacher during teaching and learning activities matches what is tested. As the test was aimed for grade 3 elementary students, the content included simple vocabulary and grammar, and asked students to start writing simple sentences. It is in line with what Rea-Dickins & Rixon (1999) and (Widiastuti, 2023) found. Most of the questions in the teacher-made-test are supplemented by the use of pictures, which are indispensable for exams for young students. Pictures make the questions more interesting

and fun, thus increasing student interest and engagement. In addition, images can help reduce the anxiety that students may feel when facing large amounts of text, by providing visual breaks that make the questions seem easier. Images also help students understand vocabulary better, provided that the item is in their first language. However, the use of images in the test papers is still monochrome, which reduces their effectiveness. For example, in question number 14 on the Indian flag, the colourless image will make it difficult for students to recognise it although the teacher added *Taj Mahal* as another clue. Ideally, all images should be uniformly coloured to provide additional context that facilitates student understanding. The teacher should also be selective in adding images to test items to keep them functional.

Some minor flaws such as punctuation and grammatical errors need to be corrected. As Waugh et al. (2021) said that the teacher should have a good knowledge and understanding of grammar and punctuation before teaching them to students, if a teacher teaches the wrong concept to students, it will have a negative impact on the process of student mastery of knowledge. Errors in concept delivery not only hinder students' understanding of the material, but also instil erroneous knowledge. Teachers' knowledge and ability to develop good tests can illustrate their level of assessment literacy. If the quality of the test still needs a lot of improvement, then this shows that the teacher's assessment literacy still needs to be improved.

Repetition of words in the stem and multiple-choice options should also be avoided as it may increase the possibility of students guessing the answers. Many think that exams for primary school students should be made simple and easy. However, this thinking needs to change. Even if there are no major mistakes in the test item writing, improvements still need to be made to improve the quality of the test itself. This will ensure that the exam not only accurately measures students' understanding but also helps them develop optimally in the learning process.

## **CONCLUSION**

The teacher-made English test items given at the middle or end of the semester as summative assessment in this research is in the domain of testing vocabulary, grammar, and simple sentences. The conclusion from the findings shows that the test is generally in accordance with the general guidelines that have been set. However, there are some aspects that require further attention like the occurrence of some errors, especially in spelling as well as punctuation, the use of unclear or out-of-context images, and the use of grammar. Specifically, the options in single and complex multiple choice should be made homogeneous. For short answer questions, the question should be ensured to have one correct answer. In essay questions, no significant issues were found. Improvements are needed to enhance the quality of the test.

This research provides a valuable contribution for teachers in producing better and more accurate tests in measuring students' abilities. However, this research is only limited to content

analysis of written test document. Students' answer sheets should also be included in order to check the validity of the questions. In addition, interviews with the teacher should be conducted to enquire about the clarity of the instructions given and to check whether key errors in the test items affect the students, for example, when the test was given whether there were students who asked questions about the meaning of the questions. Research should also consider the procedures for conducting summative examinations in school.

## REFERENCES

- Afriyanti, R. (2015). An analysis of English tests at pilot international standard school in 2010/2011 – 2011 /2012 academic year. *Teaching English Linguistics Literature Usage (TELL-US) Journal*, 1(1).
- Beach, R. (1993). *A teacher's introduction to: Reader-response theories*. NCTE.
- Benigno, V., & de Jong, J. (2016). The “Global scale of English learning objectives for young learners”: A CEFR-based inventory of descriptors. In M. Nikolov (Ed.), *Assessing young learners of English: Global and local perspectives* (Vol. 25, pp. 43–64). Springer. [https://doi.org/10.1007/978-3-319-22422-0\\_3](https://doi.org/10.1007/978-3-319-22422-0_3)
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27–40. <https://doi.org/10.3316/QRJ0902027>
- Brown, H., & Abeywickrama, P. (2019). *Language assessment: Principles and classroom practices* (3rd ed.). Pearson Education.
- Butterfield, J. (Ed.). (2015). *Fowler's dictionary of modern English usage* (4th ed.). Oxford University Press.
- Creswell, J. W. (2015). *A concise introduction to mixed methods research*. SAGE Publications.
- Cunningham, G. K. (1998). *Assessment in the classroom: Constructing and interpreting texts*. The Falmer Press.
- De Silva, R. (2021). Assessment literacy and assessment practices of teachers of English in a South-Asian context: Issues and possible washback. In B. Lanteigne, C. Coombe, & J. D. Brown (Eds.), *Challenges in language testing around the world* (pp. 433–446). Springer Singapore. [https://doi.org/10.1007/978-981-33-4232-3\\_31](https://doi.org/10.1007/978-981-33-4232-3_31)
- Erna, N., Jabu, B., & Salija, K. (2018). *An analysis of English teacher-made tests*. Makassar State University.
- Garner, B. A. (2016). *Garner's modern English usage* (4th ed.). Oxford University Press.
- Kansas Curriculum Center. (2001). *Is this a trick question?: A short guide to writing effective test questions* (B. Clay & E. Root (Eds.)). The Kansas State Department of Education.
- Lebagi, D., Sumardi, S., & Sudjoko, S. (2017). The quality of teacher-made test in EFL classroom at the elementary school and its washback in the learning. *Journal of English*

- Education, 2(2), 97–104. <https://doi.org/10.31327/jee.v2i2.289>
- Malloy, A. (2015). Seven essential considerations for assessing young learners. *Bio Techniques*, 24(1), 20–23.
- Marsevani, M. (2022). Item analysis of multiple-choice questions: An assessment of young learners. *English Review: Journal of English Education*, 10(2), 401–408. <https://doi.org/10.25134/erjee.v10i2.6241>
- Mina, R., & Wijaya, B. (2014). An analysis on the English multiple-choice item test for primary students. *Jurnal Pendidikan Dan Pembelajaran Khatulistiwa*, 3(7), 1–12.
- Purpura, J. E. (2016). Second and foreign language assessment. *The Modern Language Journal*, 100(S1), 190–208. <https://doi.org/10.1111/modl.12308>
- Pusmendik. (2019). *Panduan penilaian tes tertulis*. Kementerian Pendidikan dan Kebudayaan Republik Indonesia.
- Rea-Dickins, P., & Rixon, S. (1999). Assessment of young learners' English - Reasons and means. In S. Rixon (Ed.), *Young learners of English: Some research perspectives* (pp. 89–101). Longman.
- Sarıgöz, İ. H., & Fişne, F. (2018). English language assessment and evaluation practices in the 4th grade classes at main stream schools. *Journal of Language and Linguistic Studies*, 14(3), 380–395.
- Septi, N. K. D., Paramartha, A. G., & Wahyuni, L. G. (2020). An analysis of the quality of teacher-made multiple-choice test used as summative assessment for English subject. *Jurnal Imiah Pendidikan Dan Pembelajaran*, 4(1), 356–368.
- Stiggins, R. (1988). Revitalizing classroom assessment: The highest instructional priority. *Phi Delta Kappan*, 69, 363–368.
- Torres, J. O. (2019). Positive impact of utilizing more formative assessment over summative assessment in the EFL/ESL classroom. *Open Journal of Modern Linguistics*, 09(01), 1–11. <https://doi.org/10.4236/ojml.2019.91001>
- Waugh, D., Waugh, R., & Warner, C. (2021). *Teaching grammar, punctuation and spelling in primary schools* (4th ed.). Learning Matters.
- Widiastuti, I. A. M. (2023). *Assessing young learners* (L. Artini & I. B. Mantra (Eds.)). Unmas Press.
- Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, 70, 703–710.