

CAN STUDENT ENGLISH COMPETENCE BE MEASURED BY USING STANDARDIZED TEST?: A REFLECTION OF PREVIOUS ENGLISH EXAMINATION ON ASSESSMENT POLICY

Diterima: 29 Januari 2023

Direvisi: 04 Februari 2023

Disetujui: 30 Maret 2023

Alfian

UIN Sulthan Thaha Saifuddin Jambi

Email: alfian@uinjambi.ac.id

Abstract: Assessment is one of the important aspects in the education process. However, the assessment that relies on the examination, especially written test cannot describe the student's real competence in English especially for speaking skill. This paper presents the reflection of previous policy in the assessment in English subject. This paper argues that there is a discrepancy between the assessment practice and assessment theory which is implemented by the ministry of education. In addition, this paper also suggests several methods of assessment in English and provides the principles of language assessment and its implementation in the examination, finally this paper also recommends to accommodate school-based and external assessments.

Keywords: examination, English, competence, assessment

Abstrak: Penilaian merupakan salah satu aspek penting dalam proses pendidikan. Namun, penilaian yang bergantung pada ujian, terutama tes tertulis tidak dapat menggambarkan kompetensi nyata siswa dalam bahasa Inggris terutama untuk keterampilan berbicara. Tulisan ini menyajikan refleksi kebijakan sebelumnya dalam penilaian mata pelajaran bahasa Inggris. Tulisan ini berpendapat bahwa terdapat ketidaksesuaian praktik asesmen dan teori asesmen yang dilaksanakan oleh kementerian pendidikan. Selain itu, makalah ini juga menyarankan beberapa metode penilaian dalam bahasa Inggris dan memberikan prinsip-prinsip penilaian bahasa dan implementasinya dalam ujian, akhirnya makalah ini juga merekomendasikan untuk mengakomodasi penilaian berbasis sekolah dan eksternal.

Kata kunci: ujian, bahasa Inggris, kompetensi, penilaian

1. Introduction

Student's assessment is a critical aspect of curriculum, providing information and evidence of learning (Alfian, et al, 2022). The information from assessment could be used

to modify teaching, curricula, program which aims to improve learning (Black & William, 1998; Brookhart & Nitko, 2007; Linn and Miller, 2005). Currently, student's assessment in the form of examination is still being used in some educational institution. In Indonesia, assessment policy was previously called National Examination that has been implemented for long period of time. This policy was issued based on the Ministry of National Education regulation (MoNE) of Indonesia No. 19/2005 and Ministry of national Education and Culture (MoNEC) No. 66/2013. The main purpose of the examination policy is to measure students' achievement in the group of science and technology subject's areas at grade 6, 9, and 12. The results of the national examination were mainly used to map educational quality nationwide, enhance educational accountability and to determine student graduation, as well as to place student in the higher level of education (MoNE regulation no. 19/2005 article 66 and 68 and MoNEC regulation No. 66 chapter 2).

There were basically three subjects assessed in the group of science subject area of the examination; Indonesian language, English language, and Mathematics. However, for secondary level there were another three additional subjects, such as natural science or social science, physics, biology, sociology, etc. depending on the students major. These subjects were assessed by using a standardized test designed by the government in which the method of the test is constructed on the basis of multiple choices (BNSP, 2013). In terms of the purpose and methods used in the examination policy, there seems to be a big gap between this examination practice and assessment theory, particularly in English examination. This examination emphasized the assessment of learning or summative assessment which are utilized to evaluate the educational achievements of individuals, typically for evaluative intentions to provide information about what learning has been achieved at a certain time (Alfian et.al, 2022; Dolin et al, 2018).

Several standard competencies in English skills should be achieved by senior secondary level student who will graduate from this level. According to MoNE of Indonesia regulation No. 23/2006 and MoNEC regulation No. 69/2013, the standard competency of senior high school graduates consists of four skills; listening, speaking, reading and writing. These skills cover understanding and expressing written and oral formal and informal interpersonal and transactional discourses. This is in line with Brown (2004) who states that the language user will perform the act of four language skills. They depend on the examination underlying competence in order to accomplish the performance. Furthermore, according to Speinelli (2006) language assessment involves the ability to talk, understand, and process information. Thus, these competencies will be identified by assessing student's competencies in English to see their achievement.

From standard competency of MoNE regulation, Brown (2004) and Speinelli (2006) perspective, it is explicitly clear that English assessment should reflect the competency of student in all skills that have been learned. Thus, assessment will be relevant to the teaching and learning because what is being assessed match with what has been taught (Wu & Hornsby (2012). Furthermore, the principal or effectiveness of language assessment and method of assessment should also be considered. Thus, this will indicate student's

achievement in English as intended in examination policy which is to measure student's achievement. Until recently scarce studies investigate the method, the principle as well as the possible external examination involved in assessing students' competence. Thus, this study aims to explore the method used, principle of assessments and possible external factor in assessing students learning. This paper is initiated by introduction followed by method of the study and continue with the findings and discussion. The last section is conclusion of the study.

2. Methods

The purpose of this study is to explore the method used, principle of assessments and possible external factor in assessing students learning. Thus, this study uses the library research method. According to Bryman (2016), social researchers can benefit greatly from using library research as it provides access to previously collected and analyzed information by other researchers. This review of existing literature can help identify areas where knowledge is lacking, explore various theoretical perspectives, and gain a deeper understanding of a topic. Furthermore, Moser and Kalton (2018) assert that secondary data can be a useful option for researchers because it can be obtained quickly and at a lower cost than primary research methods. The present study relies on secondary data sources such as published books, academic journals, documents and other works. As a result, the data analysis method used is content analysis, which involves a systematic and objective examination of the content of written material or texts (Krippendorff, 2018). Content analysis can reveal patterns, themes, and other meaningful information in the literature, which can aid in developing a better comprehension of the research topic.

3. Findings and discussion

a. Method of assessment in English

Based on the data analysis, there are several methods that can be used in assessing student's competency which is basically divided into two kinds: test/examination and non-test. However, most teachers tend to use test method for assessment. For example, the method which was used in the English examinations was dominantly multiple choice which items consisting 50 questions that must be completed within 120 minutes. These 50 questions are divided into two categories, listening and reading in which 15 items for listening (understanding pictures, giving responses, understanding dialogues) and 35 items for reading (understanding written dialogues, advertisement, and reading text) (POS UN, 2013). This mean that it is only receptive skills tested while productive skills, writing and speaking which are part of English skills are not tested in the examination. The multiple-choice method as a standardized test used in English of examination is categorized into high stake standardized test since the result will "carry a serious consequences for students and teacher or educator" Marchant (2004). The examination result based on the MoNE regulation no. 19/2005 article 66 and 68 and MoNEC regulation No. 66 chapter 2, was used to determine students graduation which means that they

could pass or fail. This standardized test seems to have more disadvantages than the advantages.

Standardized test tend to emphasized segmental skill (Cognitive: mental skill (knowledge), rather than testing higher order thinking skill, creativity, and problem solving (Lerner, 2003) and Affective (attitude), and Psychomotor (skills) domains (Propham (1999). So, this standardized test “tends to be unreliable, and give incomplete and uninformative pictures of student’s achievement” (Knight, 2002. p. 107). Furthermore, standardized test indicated the discrepancy between what is tested and what is covered in the curriculum (Spinelli, 2006). In the English curriculum, it is stated that stated students need to be able to communicate which mean spoken and written in English. However, the English test in examination only covers reading and listening skills. This means that English of examination as the assessment is not relevant to the curriculum. In another words, teaching and learning is not relevance with the assessment because what is assessed did not match what has been taught (Wu & Hornsby, 2012).

Standardized test also provides the advantages. It is believed that it sets meaningful standard and acknowledges its credibility in enabling greater consistency, transferability and comparability of results across jurisdictions (Santiago et al. (2011). This assessment tend to be more time and cost efficient, requires less teacher involvement, and are useful for state wide and national administrative, policy making and reporting system (Spinelli (2006). From this perspective, it is the fact that English for segmental skills throughout Indonesia can be compared through examination. However, it cannot be used to show students achievement. In terms of cost efficient, it seems that examination 2013 spends more than Rp 644.24 billion (US\$ 66.29 million) (the Jakarta post, 2013). This means that the examination in general is not cost efficient because US\$ 66.29 million is very big for Indonesian.

From the above explanation, it is clear that there is single method with single task used in English of examination to measure student competency. Furthermore, this English of examination does not assess the whole skills that have been taught or that student should possess. Moreover, the disadvantages of standardized test used English of examination outweigh its advantages. This test result shows a little information about students’ achievement. This means that the government purpose in measuring the student’s English achievement cannot be established. Furthermore, single methods of assessment seems does not accurate to clearly indicate student competence as suggested by Brown (2004) which is supported by Linn and Miller (2005) who states that assessment seek information by using variety of procedures and products at understanding and improving learning.

Brown (2004) suggested several methods that should be considered in assessing a language learner such as, several testing (test repeatedly), a single test with multiple test task, in – class and extra class graded work, conference, journal portfolio, self-assessment, observation, and peer assessment. These methods of assessment were also recommended by the directorate of higher education, such as test, portfolio, performance, peer assessment, self-assessment, observation, etc (UIN STS Jambi, 2021). From Brown (2004) it is true that accommodating these assessment methods, the student competence could be assessed accurately. He also argues that the principle of language assessment, such as, practicality, validity, authenticity, washback, and the reliability should also be incorporated in assessing students. However, there seems to be a gap between the previous practice of English examination policy assessment to measure student achievement and language assessment theory suggested by Brown (2004).

b. Principles of language assessment and its implementation in the examination

Based on the data analysis, the directorate of higher education (2019) suggested the principle of assessment as follow:

1. Education : the assessment is aimed to improve the way of learning and achieve learning outcome.
2. Authentic: the assessment should guarantee the continuity of learning and it should reflect students' ability.
3. Objective: assessment should be in agreement between student and teachers and there is no conflict of interest between subject.
4. Accountability: The assessment in accordance with the procedure and criteria
5. Transparence: the result of the assessment can be accessed with stakeholders

As stated above that the principal or effectiveness of assessment and language assessment as suggested by the theory should be considered in assessing students learning. In line with the principle above, Brown (2004) has suggested several criteria of effective assessment, such as reliability, validity and authenticity, etc. The validity of the English in examination instrument is questionable since the test content which is part of the most important skills in the language is not included in the test items. According to Brown (2004) validity is concerning with the extent to which the test indicate the performance that matches with the course. Several validities might need to be considered in the assessment instrument. The first is content validity. Brown (2004) stated that a test is said to have content validity if the test items constitute a representative sample of language skills, structure, and language component etc., for example, speaking test should be made up of the items testing the ability of speaking. From this view; it is clear that English examination in examination is lack of content validity because the test instrument does not represent the four skills in English.

The next validity is face validity. "face validity refers to the degree to which a test looks right, and appears to measure the knowledge or abilities it claims to measure, based on the subjective judgement of the examinees who take it, the administrative personal who decide on its use, and other psychometrically unsophisticated observers" (Mousavi, 2002, p. 244). The English of examination does not measure the knowledge of writing and speaking. Therefore, it is lack of face validity. According to Brown (2004). Face validity will be high if it is well constructed, task related to the coursework (content validity). The examination may be well constructed. However, the test doesn't incorporate part of course work or lack of content validity.

Beside the validity of the assessment instrument, the reliability is also important to be considered. Reliability is "the consistency of assessment results" (Brookhart & Nitko, 2007, p. 43) . Reliable test is consistent and dependable (Brown, 2004). This means that the degree to which students result are the same when (a) Students do the same task on two or more different time (b two or more teachers mark the presentation (c) student completes two or more different but equivalent task on the same or different time (Brookhart and Nitko, 2007). According to Lin and Gronlund (1995 the reliability of standardized test is commonly high .80 and .95: frequently is above .90. Since the English is measured by using standardized test, this mean that the reliability of English is high.

Another concern about the English of examination is the authenticity. Wiggin, 1990 (as cited in Brookhart and nNitko, (2007) suggests that authentic task has the following characteristics: it requires student to utilize their knowledge to do meaningful task. It also

requires students to utilize combination of different knowledge, skill and abilities. Authentic tasks also require a high quality polished, complete, and justifiable responses, performance or products. Brown (2004) argues that an authentic language test should be as natural as possible, the topic are meaningful and real world task.

Having discussed the methods and the principle of language assessment in the theory compared to the English examination practice, it seems that the government purpose to measure student achievement in English may not be achieved based on the several reasons. The first, the principal language assessment, such as the validity of the English of examination is very low; it doesn't represent the validity of good instrument in term of the content, face and construct validity. Beside validity, the test is far from being authentic although the reliability of the score and the test items are not arguable since it is a standardized test. The second, it employed a method of assessment with single task used. This method covers very limited aspects of learning. It is a fact that writing and speaking do not cover in the examination.

Although the English of examination, according to the theory may not measure student's achievement comprehensively, the results are partially used to determine the graduation of students. This seems that there is a mismatch between the graduation and the achievement. Graduation of students are preferably determined by the achievement, while it seems that examination may partially show students achievement, accommodating teacher assessment will be very important to be considered. Student's achievement can be assessed using variety of assessment by teacher and school assessment. The following section will discuss the alternative assessment to accommodate the purpose of the assessment policy.

c. Accommodating school-based and external assessments

The practice of English assessment in examination policy seems mismatched with the theory of English assessment. The intended purpose of English language assessment to indicate student's achievement may not be established through the examination. This also mean that the use of the result of examination for graduation, mapping educational quality, and placing students in the higher level of education could be invalid. Furthermore, there is a discrepancy between what is assessed and what has been taught. In another word, the assessment is not relevant with the teaching and learning. It seems that the examination is weighed to the administrative aspects for public accountability. The government ensures that the appropriate educational standard nationwide should be achieved in order to meet the expectation of stake holders. This also means that the primary focus of the assessment on the outcome and measurement and the secondary focus are on learning (Boud, 2007). The English achievement could not be measured through examination practice; however, the government just want to show that students have been assessed.

Kivisto (2008) argues that government demands for accountability lead to reductions in autonomy and freedom in academic activities. This is certainly true for assessment as reflected in current assessment in English of examination Policy. For example, teacher role to conduct the assessment as stated in Law no. 20/2003, article 58 is restricted. The teacher role in examination is just a trainer for helping students to answer the examination items. This means that teaching and learning approach in the classroom is also changing. Teaching and learning process are not aimed to improve students learning in English , but to drill students in order to be able to answer the question on the test (Cimbricz, 2002; Paris& Urdan, 2000) and teachers

feel they are forced “to teach to test” (Spinelli, 2006). So how can students’ competency be achieved while students and teacher are busy in preparing for the exam.

Many scholars and stakeholder ask question why government keep maintaining the examination while it adopts traditional approach of assessment, assessment *of* learning, in which this approach of assessment is used for making judgement of students and for accountability as well as to make decision about student’s future program or placement (Earl, 2003). The answer to this question is that the examination policy like any other policy is “reactive rather than visionary (Meyer, et al (2010) which could mean that the main objective of conducting the examination is really for educational accountability and for determining student graduation.

Surely, assessment *of* learning like examination accommodates external assessment which relies on the standardized test. This assessment relies on the summative test which provides several disadvantages as explained above and there is no feedback to the students and teacher which means that student hardly interpret their leaning about where to go, how to go and where next to. (Hattie, 2003). Therefore, to optimize teacher role in assessing their students and improve learning, it seems that assessment *for* learning could be employed within the Indonesian context as an alternative assessment. Assessment *for* learning is ongoing, diagnostic, and formative assessment through classroom assessment in which teacher collects information or data by using varieties of resource, portfolio, observation, and test, etc. This assessment occurs frequently in the middle of learning as a formative assessment. It is interactive assessments in which teachers also provide feedback to scaffold next steps of learning which can enhance teach (Earl, 2003). This assessment is also conducted repeatedly as the continuous assessment which will improve performance and behaviour (Ridderinkhof, Ullsperger et al. 2004).

It seems that accommodating the classroom assessment or assessment *for* learning will show the greater possibility that student achievement would be measured accurately and could improve learning as well as improve their competence in English. Furthermore, variety of assessment methods could also be employed by the teacher. So, the result is very reliable to be used as the graduation criteria and other purposes as intended in the examination policy. However, the major concern about schools based assessment within the Indonesian context is that the teacher and students’ knowledge about principal of language assessment may be limited. Teacher may not have a wide knowledge about methods and principal of good language assessment. According to Sulistyono (2009) who conducted research on the national examination about the teacher opinion about examination and classroom assessment. He found that the majority of teachers (62,5%) prefers examination (external assessment) while the other teachers (37%) prefers to use teacher made test (can be categorized into school based test). The reason why most teachers prefer to have an examination or external exam is that because it is more objective and it reflects national standard and they are still questioning their knowledge about principle of language assessment. One solution that might be used to overcome this problem is by upgrading teacher`s knowledge about assessment.

4. Conclusion

The main purpose of examination which is to measure students’ achievement may not be established according to the theory of language assessment. The method of the assessment and the principle of language assessment in examination do not reflect the theory of language assessment. Furthermore, the examination policy adopt assessment *of* learning which may give

a little impact to learning and show a little information about student achievement, therefore, it is suggested to move to assessment *for* and *as* learning. These assessments would optimize teacher role in assessing students and varieties of methods in assessing students will also possible to be used so that student English competency and achievement will be accurately measured. It is recommended that teacher made assessment is solely used to determine student's achievement and graduation. On the other side examination is just used to map the standard of education which can help government to make generalization of educational quality by nationally.

References

- Alfian, A., Rasyid, M. N. A., Habibi, A., Noprival, N., & Yusuf, M. (2022). Classroom Assessment Practices of EFL Lecturers with Current Curriculum Implementation: Where Policy Meets Practice. *REiLA: Journal of Research and Innovation in Language*, 4(3), 320-334. <https://doi.org/10.31849/reila.v4i3.11005>
- Black, P., & William, D. (1998). Inside the black Box: Raising Standards through Classroom Assessment. Retrieved on Jexaminatione 5, 2013 from http://flo.flinders.edu.au/pluginfile.php/551209/mod_resource/content/1/BlackWiliam_1998_BlackBox.pdf
- Bryman, A. (2016). *Social research methods*. Oxford University Press.
- BNSP (2013). Prosedur operasi standard pengelenggaraan Ujian nasional 2013. Retrived on July, 3, 2013 from http://bsnp-indonesia.org/id/bsnp/wp-content/uploads/2013/02/0020_-Peraturan-POS-EXAMINATION-SMPSMA-SMK-dan-EXAMINATIONPK-Tahexamination-2013-4-Feb-2013_-Edit-BSNP.pdf
- Boud, D. (2007). Reframing assessment as if learning were important. In D. Boud & N. Falchikov (Eds.), *Rethinking Assessment in Higher Education* (pp. 14-25). London: Routledge.
- Brown, James Dean. (1996). *Testing in Language Programs*. Upper Saddle River, New Jersey: Prentice-Hall Regents.
- Center for policy option. (1993). *Outcomes – based accoexaminationtability: policy issues and options for students with disabilities*. Rocville, MD:Westat
- Cimbricz, S. (2002). State-mandated testing and teachers' beliefs and practice [Electronic version]. *Education Policy Analysis Archives*, 10(2), 1-22.
- Dolin, J., Black, P., Harlen, W., & Tiberghien, A. (2018). Exploring relations between formative and summative assessment. *Transforming assessment: Through an interplay between practice, research and policy*, 53-80.
- Earl, L. (2003). *Assessment as learning: using classroom Assessment to Maximise Students Learning*. Thousand Oaks, CA: Corwin Press.

- Falchikov, N. (2005). *Improving assessment through student involvement*. London and New York: Routledge.
- Gronlexaminationd, Norman E. (1998). *Assessment of student achievement*. Boston: Allyn and Bacon
- Hattie, J. (2003). *Formative and Summative Interpretations of Assessment Information*. New Zealand: The Examinationiversity of Auckland
- Kivistö, J. (2008). An assessment of agency theory as a framework for the government–examinationiversity relationship. *Journal of Higher Education Policy and Management*, 30(4), 339-350.
- Knight, P. (2002). The Achilles' heel of quality: The assessment of student learning. *Quality in Higher Education*, 8(1), 107-115.
- Krippendorff, K. (2018). *Content analysis: An introduction to its methodology*. Sage publications.
- Kvale, S. (2007). Contradictions of assessment for learning in institutions of higher learning. In D. Boud & N. Falchikov (Eds.), *Rethinking assessment in higher education* (pp. 57-71). London: Routledge.
- Lerner, J. (2003). *Learning disabilities: theories, diagnosis, and teaching strategies*. Boston: Houghton Mifflin
- Linn, J.E. & Gronlexaminationd, N. E . (1998). *Measurement and Assessment in Teaching*. New Jersey: Prentice Hall
- Linn, R.L. & Miller, M.D. (2005). *Measurement and assessment in teaching*. New Jersey, USA: Pearson Prentice Hall
- Meyer, L., Davidson, S., McKenzie, L., Rees, M., Anderson, H., Fletcher, R., & Johnston, P. (2010). An investigation of tertiary assessment policy and practice: Alignment and contradictions. *Higher Education Quarterly*, 64(3), 331-350.
- Moser, C. A., & Kalton, G. (2018). *Survey methods in social investigation*. Routledge
- Mousavi, Seyyed Abbas. (2002). *An encyclopedic dictionary of language testing*. Taiwan: Texaminationg Hua Book Company.
- Nitko, A. J., & Brookhart, S. M. (2007). *Educational Aessment of Students*. New Jersey: Pearson Education Inc.
- Paris, S. G., & Urdan, T. (2000). Policies and practices of high-stakes testing that influence teachers and schools [Electronic version]. *Issues in Education*, 6(1), 83.
- Permendikbud (MoNEC regulation) No. 66 tahexamination 2013 (2013). Standard Penilaian Retrieved on July 3, 2013 from <http://bsnp-indonesia.org/id/?p=1239>

- Permendikbud (MoNE regulation) No. 19 tahexamination 2005 (2005). Standard nasional pendidikan. Retrieved on July 3, 2013 from <http://www.scribd.com/doc/54107157/PP-No-19-Tahexamination-2005>
- Permendikbud (MoNEC regulation) No. 69 tahexamination 2013. (2013). Kompetensi dasar dan struktur kuikulum SMA-MA. Retrieved on July 3, 2013 from <http://bsnp-indonesia.org/id/?p=1239>
- Propham W. J. (1999). Where Large Scale educational assessment is heading and why it shouldn't be. *Educational Measurement: issue and practice*. 13-17
- Salvia, J & Ysseldyke, J. E. (2004). *Assessment in special and inclusive education*. Boston: Houghton Mifflin.
- Santiago, P, Donaldson, G, Herma, J and Shrewbridge, C. (2011). OECD Reviews of Evaluation and Assessment in Education: Australia, OECD
- Spinelli, Cathleen G. (2006). *Classroom assessment for students in special and general education*. New Jersey: Prentice Hall.
- The Jakarta Post (2013). House calls for Exams annualment. Retrieved on 25 Jexaminatione, 2013 from <http://www.thejakartapost.com/news/2013/04/27/house-calls-exams-annulment.html>
- UIN STS Jambi. (2021). Pedoman penilaian pembelajaran. LPM UIN Sulthan Thaha Saifuddin Jambi
- Wu , M & Hornsby, D. (2012). Inappropriate uses of NAPLAN results. in Say no to NAPLAN. Retrieved on Jexaminatione1, 2013 form <http://www.literacyeducators.com.au/images/pdf/naplan/naplan-papers-set-1.pdf>