

Computer-Based Test and Paper-Based Test as English Language Assessment in Indonesian Junior High Schools

Henry Solekhah, Flinders University, Australia

Asian Conference on Education & International Development 2018
Official Conference Proceedings

Abstract

This paper explains and critiques the implementation of CBT (Computer-Based Test) and PBT (Paper-Based Test) as English language assessment in Indonesian Junior High Schools. The policy analysis was done by scrutinising two regulations of Badan Nasional Standar Pendidikan (Indonesian Bureau of Standardised Education): BSNP: 0075/SDAR/BSNP/XII/2016 about the contents of National Examination, and BSNP: 0043/P/BSNP/2017 about the national examination procedures. The comparison to the validity and reliability of English language assessment was also done based on the implementation in 2017. There are some findings: 1) The regulation reduced the high-stakes of national examination; 2) 4,2 million examinees did two different administration procedures in the examination: 11,096 schools (1,349,744 students) used CBT, while 2,855,633 students in 45,092 schools used PBT ; 3) Sixty percent of 11,096 schools could do CBT independently whereas the others should take test on other schools, 4) The content and construct validity of the English testing was challenged by the fact that the listening and speaking skills were not assessed in both CBT and PBT, the use of multiple choice could not accommodate students' higher-order thinking, and the educational gaps among Indonesian regions; 5) The reliability of this assessment was also reduced due to the different forms of administration, technological barriers, and test schedules causing different psychological impact on the test takers; 6) Although there were limitations of these policies implementation, the Indonesian government was optimistic to increase the quality and quantity of CBT use in the national examination to improve the accountability of Indonesian education.

Keywords: Computer-Based Test, Paper-Based Test, Indonesian English Examination, English Language Assessment

iafor

The International Academic Forum
www.iafor.org

Introduction

English is a foreign language mandatory taught in Indonesia. To ensure that the educational quality in Indonesia is improved, the government conducts the national examination annually (Idrus, 2012), in which students' English competency is also assessed. The type of examination is considered summative assessment because is done at the end of the course to know how good or how bad the students' achievement is compared to the expected learning outcomes (Brady & Kennedy, 2012). Although the common response to test is mostly negative (Lamprianou & Athanasou, 2009); the importance of knowing the students' learning, the needs of society and economy, as well as the public funding to education enforce the necessity of education' accountability (Brookhart, 2011). This article will explain about the implementation of English examination in junior high schools based on the demography of test takers, the issue related to the validity and reliability of Computer-Based Test and Paper-Based test used as the media of examination, and the recommendation for better implementation in the future.

Demography of Test Takers

In implementing the national examination, the ministry of education and culture in Indonesia has a department called *Badan Standar Nasional Pendidikan* (Indonesian Bureau of Standardised Education), that focuses on the standardised education. The implementation of national examination is based on two regulations 0075/SDAR/BSNP/XII/2016 about the contents of National Examination, and BSNP: 0043/P/BSNP/2017 about the national examination procedures. These regulations have some impacts on the implementation: compared to the previous national examination, the current examination is less high-stakes; and there are two types of media used in test administration (Computer-Based Test and Paper-Based test). High stakes mean that the result determines the students' future, such as 'fail' or 'pass' (Plake, 2011), less high stake meaning that the current examination is not the only factor to make the students graduate because the other factors such as schools' achievement and portfolios are considered (Nugroho, (2017). Demographically, in 2017 there are 20% of schools with 1,349,744 students used Computer-Based Test (CBT), and 68% of school with 2,855,633 students did the exam in Paper-Based Test (Pengembangan, 2017).

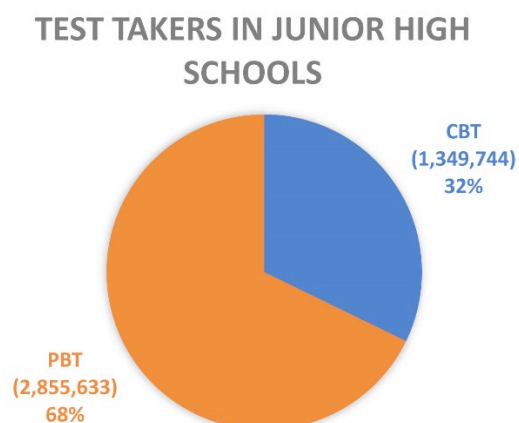


Figure 1. The percentage of test takers in junior high schools

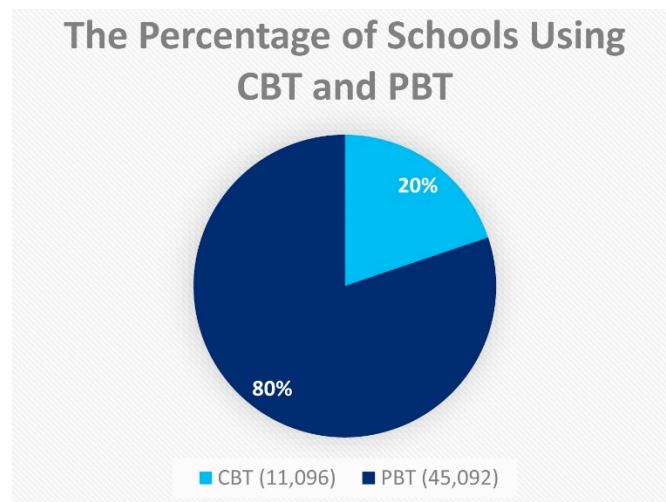


Figure 2. The percentage of schools using the CBT and PBT

The figure 3 below illustrates the spread of CBT in some areas in Indonesia. It could be seen that most the area covered with blue ink are Java island, Sumatera island has some blue spots, while the other island have significantly fewer spots. It indicates that the is significant difference in the administration of national examination.



Figure 3. The spread of CBT for junior high school (source: Pengembangan, 2017)

Validity

“Educational achievement can be defined as the extent to which specified objectives are accomplished by individual students” (Payne, 1974, p. 29). In assessing students’ achievement, it is very important to have valid and reliable measurement. Validity refers to how accurate the assessment reflects learning objectives being assessed (Brady & Kennedy, 2009). Among different types of validity, the construct validity and content validity will be discussed in this article.

Construct Validity

Construct validity is related to the fairness, relevance, and meaning of the assessment to provide accurate information (Brady & Kennedy, 2009). It relates to “Who, what, and how are we testing? What system will we use to score it?” (O'Sullivan & Weir, 2011, p. 23). Since construct validity brings the meaning of assessment psychologically (Lyman, 1978), the assessment should not favour one group of people over the others. Assessment should be fair meaning that it gives equal opportunity to the candidate to demonstrate their abilities (Isaacs, Zara, Herbert, Coombs, & Smith, 2013). However, the standardized test does not measure creativity and unfairly treat a group over another (Lyman, 1978; Gordon, & Rajagopalan, 2016).

Hadi (2014) states that national examination is not suitable to evaluate education in Indonesian. Although the passing grade of English is low (score 55 or 28 correct answers of 50 questions), the different conditions between metropolitan cities and remote area challenge the construct validity. It is undeniable that there is a big educational gap among areas in Indonesia, such as the availability of teachers, the school facilities, and the access to have information (Vito, Krisnani, & Resnawaty, 2015). These gaps, consequently raises the protests not only from students and teachers in rural areas, but also from educational experts. The government should be more thoughtful in the policy of national examination (Muntholi'ah, 2013).

Content Validity

Content validity relates to how well the measurement links to the curriculum (Brady & Kennedy, 2009). The items of English examination are 50 multiple choices assessing reading and writing. This type of test is considered the most efficient compared to other forms although there are some limitations. Multiple choice has many advantages such as broader content, quick and objective marking, but restricted in written form and ineffective for problem solving (Brady & Kennedy, 2009). “Multiple choice items are efficient and reliably scored, measure many cognitive characteristics validly, but difficult to measure higher order thinking and easier to cheat “(Geisinger, 2011, p. 241).

To comprehend reading, vocabulary and background knowledge are essential (Kane, 2011; Unruh, 2017). It enables children to associate words and context of topics (Blachowicz & Fisher, 2010; Vacca et al., 2012; Winch, Johnston, March, Ljungdahl, & Holliday, 2014). Therefore, word comprehension influences language comprehension (Gustafson, Samuelsson, Johansson, & Wallmann, 2013). Thus, vocabulary significantly influences comprehension; the more words students know, the easier for them to comprehend (Bayetto, 2013, 2014).

The figure 4. below is the example of English question in 2016, it is testing students' vocabulary in biology, specifically about the description of adult butterflies. By doing this question, it expected that the students can connect their prior knowledge to choose correct vocabularies and complete the paragraph. In question number 41, the students should choose a word that closely related to *nectar*. The clue that can help them is the next sentence that *adult butterflies* cannot chew solids. The correct answer would be (d) “liquids”. However, the question number 42 is more difficult to know what “proboscis” is. In writing the test, the items should meet the leaning outcomes

and be free from irrelevant materials (Brady & Kennedy, 2009). It should be noted that English has many kinds of genres, but the items made should consider that the focus should be more about assessing the skill in English, not about natural science.

For questions 41 to 43 choose the correct words to complete the text below.

Adult butterflies can only feed on ... (41), usually nectar. Their mouth parts are modified to enable them to drink, but they cannot chew solids. A proboscis, which functions ... (42) a drinking straw, stays curled up under the butterfly's chin until it finds a source of nectar or other liquid nutrition. It then opens the long, tubular structure and ... (43) up the nectar.

41. A. fruits
 B. mud
 C. leaves
 D. liquids

Figure 4. Completing the paragraph (Pendidikan, 2016)

Another example of testing vocabulary is by choosing correct synonym based on the text. In the figure 5, it is expected that the students know the best word that have the closed meaning. The students that are familiar with this expression can answer this type of question.

27. "..., I change my mind." (Paragraph 2)
The underlined word is closest in meaning to
- A. decision
 - B. ambition
 - C. thought
 - D. suggestion

Figure 5. Completing the paragraph (Pendidikan, 2016)

Another skill tested is the students should make an inference from the statement given. In the figure 6, the students should connect the phrase such as *high goals, ideas, inspires, the year ahead*, with the options provided. The possible 'correct' answer would be *keep on success in the future*.

50. "May the high goals and ideals which led to this special honour continue to inspire you in the years ahead."
We can infer from this statement that the writer hopes Davish
- A. stay humble and calm
 - B. always be happy
 - C. keep on success in the future
 - D. be a better person

Figure 6. Making inference Completing the paragraph (Balitbang, 2017)

Academic language contexts are typically assessed through “measures of reading and writing ability” (Romhild & Bovaird, 2011, p. 61). It could be seen that the test items are academic but not socially contextual because there is no speaking and listening. Further, Darling-Hammond, Ancess and Falk (1995, as cited in Brady & Kennedy, 2009) state that the standardized testing requires single correct answer, emphasizes teacher focus more on teaching to do the test, and narrowly promote the view of curriculum. In multiple choice, the students have limitation to choose only one ‘correct’ answer, while there would be possibility that they may have different ideas in interpreting the text.

Another aspect in content validity is the cognitive domains. The test items are based on cognitive domain although it does not guarantee that all test takers do the test by thinking critically. *Knowledge* emphasises the remembering (recall, ideas, material, or phenomena); *application* requires comprehension to apply the knowledge; *analysis* emphasises the breakdown of conveying the meaning and making the conclusion (Bloom, 1994). Nevertheless, the national examination influences English teaching practices in Indonesia are more about passing the standard or examination rather than supporting the higher order thinking (Saukah & Cahyono, 2015). Hawanti (2011, as cited in Zein, 2017, p. 57) also states that test-oriented learning limits the English teaching to focus on doing the test rather than communicative competence. Therefore, the content of English examination should be more emphasized to assess both academic and communication and should be motivate positively in the teaching and learning process.

Reliability

While validity refers to the extent the test to measure the intended objective. Reliability refers to the extent of consistently measure performance (Brady & Kennedy, 2012). However, “a test may be highly reliable without being able to do any specific task well” (Lyman, 1978, p. 7), because high reliability is important but may not be sufficient for good validity. The assessment is considered reliable if there are repetition of the same test, the equivalent forms of a test, comparable parts of a test, and item data (Payne, 1974).

There are “intrinsic problems in converting a paper-and-pencil test to a CBT” because of the production and administration (Stout, 2002, p. 103). Before the implementation of Computer-Based Test, there is only one type of test administration (paper-based test), meaning that all students do the same method. The introduction of new media (CBT) influences the reliability because testing conditions can make a great deal of differences in test results (Lyman, 1978). The paper-based is less secure in the distribution and scoring, less efficient, but has less problem in technology. Computer-based test can prevent cheating, disadvantages are the slow network, the electricity problem, and the login problem (Harmiyuni & Sailan, 2016). While paper-based is conducted once a day, computer-based is done three times because there are more students than the number of computers. In doing CBT, students should wait for hours to do the test and it causes their psychological aspects. Another problem is some problem electricity, internet connectivity. The different types of administration also bring more gap among schools and areas. For example, among 54 schools in Kendari, only 10 that can use CBT. There are 3.285 students do paper test (Putsanra, 2017).

Related to this problem, the government has commitment to increase the number of CBT in the future.

Recommendation

The national examination has improved to use more advanced technology in test administration. To improve the construct validity, the content validity, and reliability, the educational gaps should be minimized. The government should improve the quality of English language education especially in rural areas by increasing the number of teachers, provide more accessible access to information, and develop educational infrastructure. Not only that, the government should also emphasize more on developing listening and speaking skills in the examination because the purposes of learning language are both academic and communication. Since the computer-based test needs more technical support, the government should ensure that electricity, internet connection, and the availability of computers are sufficient for all test takers.

Conclusion

The English national examination has the problem with construct validity due to the wide educational gap among areas, while the content validity shows that the test items only assess the reading and writing. Two kinds of administration also have impact in the reliability that makes the gap between metropolitan areas and rural areas become more obvious. The improvement in the future is highly needed, not only to support the learning outcomes, but also to bring fairness and equity in education.

References

- Bayetto, A. (2013). *Vocabulary*. Canberra: Australian Primary Principals' Association.
- Bayetto, A. (2014). *Comprehension*. Canberra: Australian Primary Principals' Association.
- Blachowisz, C., & Fisher, P. J. (2010). *Teaching Vocabulary in All Classrooms* (4th ed.). Boston: Pearson Education, Inc.
- Bloom, B. S. (1994a). Excerpts from the "Taxonomy of Educational Objectives, The Classification of Educational Goals, Handbook I: Cognitive Domain". In L. W. Anderson & L. A. Sosniak (Eds.), *Bloom's Taxonomy: A Forty-Year Retrospective* (93 ed., Vol. 2, pp. 9-27). Chicago: National Society for the Study of Education.
- Brady, L., & Kennedy, K. (2012). *Assessment and Reporting: Celebrating Student Achievement* (4th ed.). Frenchs Forest: Pearson Australia.
- Brady, L., & Kennedy, K. (2009). *Celebrating Student Achievement: Assessment and Reporting* (3rd ed.). Frenchs Forest: Pearson Education Australia.
- Brookhart, S. M. (2011). Educational Assessment Knowledge and Skills for Teachers. *Educational Measurement: Issues and Practice*, 30(1), 3-12. doi:10.1111/j.1745-3992.2010.00195.x
- Geisinger, K. F. (2011). The Future of High-Stakes Testing in Education. In J. A. Bovaird, K. F. Geisinger, & C. W. Buckendahl (Eds.), *High-Stakes Testing in Education: Science and Practice in K-12 Settings* (pp. 231-246). Washington: American Psychological Association.
- Gordon, Edmund W. & Rajagopalan, Kavitha (2016). *The Testing and Learning Revolution*. Macmillan: Palgrave.
- Gustafson, S., Samuelsson, C., Johansson, E., & Wallmann, J. (2013). How Simple is the Simple View of Reading? *Scandinavian Journal of Educational Research*, 57(3), 292-308. doi:10.1080/00313831.2012.656279
- Hadi, S. (2014). Ujian nasional dalam tinjauan kritis filsafat pendidikan pragmatisme [National examination in the critical perspective of the pragmatism of educational philosophy]. *Al-Adzka: Jurnal Ilmiah Pendidikan Guru Madrasah Ibtidaiyah*, 4(1), 283-294. doi:10.18592/aladzkapgmi.v4i2.139.g49
- Harmiyuni, & Sailan, M. (2016). Persepsi siswa tentang pelaksanaan ujian nasional berbasis komputer di smk komputer mutiara ilmu makassar [The students' perceptions about the implementation of computer-based test in mutiara ilmu makassar vocational school]. *Jurnal Tomalebbi*, III(2), 36-42.

- Idrus, M. (2012). Mutu pendidikan dan pemerataan pendidikan di daerah [The educational quality and its spread in the rural areas]. *Psikopedagogial*, 1(2), 1-10.
- Isaacs, T., Zara, C., Herbert, G., Coombs, S. J., & Smith, C. (2013). *Key Concepts in Educational Assessment*. London: SAGE Publications Ltd.
- Kane, S. (2011). *Learning and Literacy in the Content Areas*. Arizona: Holcomb Hathaway, Publishers, Inc.
- Kisi-Kisi UN Tahun Pelajaran 2016/2017,(2016).
- Lamprianou, I., & Athanasou, A. J. (2009). *A Teacher's Guide to Educational Assessment*. Rotterdam: Sense Publisher.
- Lyman, H. B. (1978). *Test Scores and What They Mean* (3rd ed.). Englewood Cliffs, N.J.: Prentice-Hall.
- Muntholi'ah. (2013). Ujian nasional, dulu, kini dan yang akan datang: Tinjauan normative [The national examination, in the past, the present, and the future: The normative perspectives]. *Nadwa: Jurnal Pendidikan Islam*, 7(1), 161-180.
- Nugroho, B. P. (2017, 24 March 2017). Mendikbud: Sekarang USBN Penentu Kelulusan, UN Malah Tidak. *detikNews*. Retrieved from <https://news.detik.com/wawancara/d-3456401/mendikbud- sekarang-usbn-penentu-kelulusan-un-malah-tidak>
- O'Sullivan, B., & Weir, C. J. (2011). Test Development and Validation. In B. O'Sullivan (Ed.), *Language Testing : Theories and Practices* (pp. 13-28). Basingstoke: Palgrave Macmillan.
- Payne, D. A. (1974). *The Assessment of Learning : Cognitive and Affective*. Lexington: D.C. Heath and Company.
- Plake, B. S. (2011). Current State of High-Stakes Testing in Education. In J. A. Bovaird, K. F. Geisinger, & C. W. Buckendahl (Eds.), *High-Stakes Testing in Education: Science and Practice in K–12 Settings* (pp. 11-24). Washington: American Psychological Association
- Pendidikan, P. P. (2016). *Ujian Nasional Bahasa Inggris Tahun 2015/2016 [English Examination 2015/2016]*. Jakarta: Kementerian Pendidikan dan Kebudayaan
- Pendidikan, P. P. (2017). *Ujian Nasional Bahasa Inggris Tahun 2016/2017 [English Examination 2016/2017]* . Jakarta: Kementerian Pendidikan dan Kebudayaan
- Pengembangan, P. P. P. B. P. d., & Pendidikan, B. S. N. (2017). *Ujian Nasional Bahasa Inggris SMP/MTs Tahun Pelajaran 2015/2016 [English Examination in Junior High School 2015/2016]*. Jakarta: Kementerian Pendidikan dan Kebudayaan.

Prabowo, D. (2017, 2 May 2017). Kemendikbud: 4,2 Juta Pelajar Ikuti UN SMP. *Kompas*. Retrieved from <http://nasional.kompas.com/read/2017/05/02/12373941/kemendikbud.4.2.juta.pelajar.ikuti.un.smp>.

Prosedur Operasional Standar Penyelenggaraan Ujian Nasional Tahun Pelajaran 2016/2017, (2017).

Putsanra, D. V. (2017, 2 May 2017). Pelaksanaan UN 2017 di Indonesia Timur Tak Semua UNBK. *tirto.id*. Retrieved from <https://tirto.id/pelaksanaan-un-2017-di-indonesia-timur-tak-semua-unbk-cnSQ>

Romhild, A., & Bovaird, J. A. (2011). Assessing English Language Skills in Education: Implications for Students with Limited English Proficiency. In J. A. Bovaird, K. F. Geisinger, & C. W. Buckendahl (Eds.), *High-Stakes Testing in Education: Science and Practice in K–12 Settings* (pp. 59– 72). Washington: American Psychological Association.

Saukah, A., & Cahyono, A. E. (2015). Ujian Nasional di Indonesia dan Implikasinya terhadap Pembelajaran Bahasa Inggris [National Exam in Indonesia and Its Implication to the Teaching and Learning of English]. *Jurnal Penelitian dan Evaluasi Pendidikan*, 19(2), 243-255.

Stout, W. (2002). Test Models for Traditional and Complex CBT. In C. N. Mills, M. T. Potenza, J. J. Fremer, & W. C. Ward (Eds.), *Computer-Based Testing : Building the Foundation for Future Assessments* (pp. 103-111). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Unruh, S. (2017). *Assessment and Intervention for English Language Learners Translating Research into Practice*: Cham : Springer International Publishing : Imprint: Springer.

Vacca, J. A. L., Vacca, R. T., Gove, M. K., Burkey, L. C., Lenhart, L. A., & McKeon, C. A. (2012). *Reading and Learning to Read* (8th ed.). Boston: Pearson Education, Inc.

Vito, B., Krisnani, H., & Resnawaty, R. (2015). Kesenjangan pendidikan di desa dan di kota [Educational gap between the villages and in the city]. *Prosiding KS: Riset & PKM*, 2(2), 247-251.

Winch, G., Johnston, R. R., March, P., Ljungdahl, L., & Holliday, M. (2014). *Literacy: Reading, Writing and Children's Literature* (G. Winch Ed. 5th ed.). Victoria: Oxford University Press.

Zein, M. S. (2017). Elementary English education in Indonesia: Policy developments, current practices, and future prospects: How has Indonesia coped with the demand for teaching English in schools? *English Today*, 33(1), 53-59.
doi:10.1017/S0266078416000407

Contact email: heny.solekhah@gmail.com, sole0014@flinders.edu.au