

Developing a Contextual English Proficiency Test: The Case of English Test for Islamic Community (ETIC)

Siti Nurul Azkiyah¹ and Abdurrosyid¹

¹Faculty of Educational Sciences, UIN Syarif Hidayatullah Jakarta, Indonesia

Keywords: Contextual test; English proficiency test; ETIC; test development

Abstract: Assessment has been widely recognized to play a very important role in education, one of which is assessment of language proficiency. For the context of English, TOEFL, IELTS and TOEIC have been widely used. However, there have been criticisms towards these types of test in terms of the dominance of English native culture and the cost of taking the test. Therefore, this study is intended to develop an English proficiency test, which is contextual for Indonesia especially Islamic community as the largest one in Indonesia. The test is named ETIC (English Test for Islamic Community), which development strictly follows the principles and the stages of an international standardized test. Three domains namely academic, Islam, and Indonesia become the main areas for the content of the test while the test specification is built based on a thorough study of competencies tested in TOEFL, IELTS as well as those mentioned in CEFR. The results of the study are expected not only to produce a unique and contextual English proficiency test but also contribute to the development of intellectual property rights of UIN Syarif Hidayatullah Jakarta as the locus of the study.

1 INTRODUCTION

It is widely believed that assessment has a very important role in education. A large number of studies have consistently indicated that assessment is essential not only for students but also for teachers and institutions (e.g. Alduais, 2012; Creemers & Kyriakides, 2008; Creemers, 1994; Hughes, 1989). Furthermore, Creemers, and Kyriakides (2008) argue that assessment is vital for teachers in providing information on what teachers should do in order to improve their teaching and learning activities. In other words, the results or data gathered from assessment should enable teachers identify the high and the low achievers in their classrooms, to analyze the weaknesses and the strengths of their teaching methods, as well as to underline the materials that need to be retaught. These are considered as necessary data for teachers to decide improvement strategies of their teaching and learning strategies.

There have been different types of assessment, one of which is proficiency test. According to some

experts such as Bachman and Palmer (1996), Carr (2011), Hughes (1989), and Kopriva (2008), proficiency test is expected to show somebody's overall proficiency in a language regardless of his/her educational background. Some examples of proficiency test in language include TOEFL (Test of English as a Foreign Language), IELTS (International English Language Testing System), TOEIC (Test of English for International Communication) etc. These types of tests are used not only for academic purposes but also for other purposes such as business and other general purposes.

The continuity of English proficiency test development is highly demanded since there is no such perfect test that can work well for all times and contexts. This view is supported by Kim (2018) who exhibits the existence of bias in English proficiency test in Seoul National University. Some experts also consider that self-developed English proficiency tests can do justice for the test takers in the form of accommodating their needs and local values (Franco & Galvis, 2012) and also being likely less biased

because all know that the tests such as TOEFL and IELTS are the tests that belong to the native speakers of English which favor the English native contexts. Also, the continual development of English proficiency test will avoid washback for English learning (Thaidan, 2015, eventhough it is not entirely).

The issue that most English tests are too “Anglobalized” are justified as it can be seen from the test framework, the values and ideologies in all parts of the test (Thomas & Breidlid, 2015). The tests have been criticized to be biased, containing values and contexts of English native speakers. Even an English test developed in non-English speaking countries like Indonesia has been found to use much more English culture rather than the source (Indonesia) culture, at 46% and 17% respectively (Azkiyah & Setiono, 2018). This is likely due to the fact that the concept of native speakers’ competence has been widely applied prior to the emergence of communicative competence (Chinh, 2013).

However, many studies have found that familiarity is a crucial aspect effecting students’ performance on the test. Sabatin (2012), for instance, mentions that the domination of the English native speakers may affect test validity due to test takers’ lack of familiarity. Similarly, Sosa (2012) considers that containing too many cultural contents may throw off some students and thus is not fair. This is because individual contexts, class, race, gender as well as cultural experience play roles in reading: readers do not passively receive texts (Tomlinson, 2010). Finally, another crucial problem is the expensive price of TOEFL or IELTS test, which is similar to a semester-tuition fee in many universities in Indonesia.

Referring to the importance of English, UIN Syarif Hidayatullah Jakarta as the context of the study, regulates that all students have to achieve a minimum standard of proficiency in both English and Arabic. Considering some arguments on the necessary to accomodate the local values and needs of the students in the test and the expensive price of the aforementioned English proficiency tests, through the language center, the university decides to develop its own test, which then is named English Test for Islamic Community (ETIC).

Therefore, this paper is intended to describe the stages of developing ETIC, the unique characteristics of ETIC as well as its items quality. In line with these objectives, three questions are raised as follows;

1) What are the stages of developing ETIC?

- 2) What specific characteristics should be attached to ETIC to make it unique and relevant to the context of Indonesia and Islamic University?
3) How is the quality of items in ETIC?

It is expected that the results of the study do not only produce a unique English proficiency test but also contribute to the development of intellectual property rights of UIN Syarif Hidayatullah Jakarta.

2 METHOD

2.1 Research Design

Concerning the design, this research will use both qualitative and quantitative methods. The qualitative approach is used to address question 1 and 2 while the quantitative one is employed to answer question 3. Firstly, an analysis of the specification of TOEFL and IELTS was conducted as a benchmark in developing the items in ETIC. This analysis is important to develop the unique specification of ETIC, which then will be used to develop the items of ETIC. The quantitative approach was then used to analyze the test items of ETIC though a pilot study intended to analyze the validity, reliability, difficulty level, and item discrimination of the test items. This analysis is important in order to make sure that every single item in the test meets the requirements of a quality test.

2.2 Study Subject

This study used both human and non-human as the subjects: the non-human ones were the TOEFL and IELTS tests while the human participants of the study consisted of the developers of ETIC and students of UIN Syarif Hidayatullah Jakarta. In this study, the team who developed ETIC consisted of 8 lecturers of English from both the Education and Literature Departments of UIN Syarif Hidayatullah Jakarta and 300 students who voluntarily attended the pilot study intended to examine the quality of the items included in ETIC.

2.3 Data Analysis

Regarding the first and the second research questions, which were the analysis of the specification of TOEFL and IELTS and development of ETIC specification, two TOEFL books and two IELTS books were analyzed to understand the specification of both types of well-

known English proficiency tests. This analysis consisted of the analysis for Listening, Structure and Written Expression and Reading. In answering the first and the second research questions, descriptive and analytical approach were used to understand the specification of TOEFL and IELTS.

For the third research question, upon the completion of the items of ETIC, the responses of students who joined the pilot study were analyzed using SPSS to analyze the validity, reliability, difficulty level, and item discrimination of ETIC.

3 FINDINGS AND DISCUSSION

3.1 Steps in Developing ETIC

This study was a long study since it involved various activities starting from the development of the test specification to the pilot study and the revision of the items. The details of each stage are presented below.

The Development of Test Specification. The development of test specification was started in March 2016, in which following the structure of TOEFL, the team was divided into three namely listening, structure and written expression, and reading comprehension. Each team analyzed the competencies examined in both IELTS and TOEFL, the lengths of the text, the type of questions, the topics of the texts, the number of questions and the way the questions was raised.

The Development of the Items. Another activity in the study which also took time was the development of the items. Each person in the team was assigned to develop 50 – 85 items so that there would be much more items than what were in the real package of items in ETIC. This activity took place from April up to August 2016 since constructing an item is very difficult.

The review and the revision of the Items. After each member of the team finished constructing the items, the draft was sent to reviewers to the first validation of the content. The results of the review was sent to the corresponding member of the team. This activity took place in July - September 2016.

The Pilot Study of the Items. After revising the items based on the feedback of the reviewers, each member of the team worked to revise the items. When the revision finished, the next big agenda was the pilot study conducted in September 2016.

The Analysis of the Items. The analysis of the item was conducted after the pilot study. The data were firstly inputted in SPSS and analyzed both

descriptively and statistically. The descriptive analysis dealt with the frequency of each response for each item. This analysis was important in order to understand the distribution of responses for each item. The statistical analysis concerned the reliability, validity and the difficulty level of each item. This activity was conducted to in September 2016.

The Revision of the Items. The second round of revision was scheduled in this study to revise the items based on the pilot study. It was recognized that students did not select certain answers offered in some items. In addition, it was also noticed that some items were both too easy and too difficult, and therefore they should be improved. The revision was scheduled in October 2016.

Applying for the Intellectual Rights Property. The last activity in the study was applying for the Intellectual Right Property to the Ministry of Law and Human Rights. The document for this Intellectual Property Right was sent to the Ministry in October 20, 2016.

3.2 The Specification and the Uniqueness of ETIC

As previously mentioned, in designing the specification of ETIC, analysis of skills examined in TOEFL (paper-based) and IELTS was conducted, the results of which served as the framework to develop the specification of ETIC. In principle, the development of ETIC considered the context of Indonesia, Islam, and academic sphere, which became the main theme of the content of ETIC. In measuring the results, ETIC adhered CEFR (The Common European Framework of Reference for Languages) which has been widely used as a reference in testing language proficiency.

Regarding the uniqueness of ETIC, different from TOEFL and IELTS, ETIC has several unique characteristics. The first is the topics or theme embedded in the test. Islam, Indonesia, and academic are the core themes included in the test. The topic on Islam as of the three topics in developing the test content of ETIC is a means of ensuring justice and fairness for Muslim when taking English test covering general knowledge and habit of Muslim around the world ranging from clothing culture and rules to architecture styles dominating the Islam world. In addition, as the second largest religion in the world with 1.6 billion adherents or 23 percent of global population as of 2010 (Chappell, 2015). The test content also proves the very existence of linguistic

interplay between English along with its native speakers and Islamic civilization (Abdurrosyid, 2017). The theme of Indonesia is pivotal to be included in the test since the country is a with the largest Islamic population in the world with 222 million adherents (“Top 10 largest”, 2018) as well as the locus where ETIC is designed and developed. The academic knowledge and information is the third main content of ETIC because the test also intended to accommodate academic test takers in general besides Muslim communities. The second is the style of constructing the questions. Instead of constructing questions, ETIC uses statements to start the items included in the test. The questioning style is used to expose the test takers to more various ways of testing their English proficiency besides common styles of questioning in English that mostly using, to be, auxiliary verbs and W-H question.

3.3 The Quality of the Items

The analysis of the item was conducted after the pilot study. The data were firstly inputted in SPSS and analyzed both descriptively and statistically. The descriptive analysis dealt with the frequency of each response for each item. This analysis was important in order to understand the distribution of responses for each item. The statistical analysis concerned the reliability, validity and the difficulty level of each item.

With respect to the validity analysis, we found some items should be improved because the coefficient of their (corrected item) total correlation was below 3. Another quantitative analysis conducted in this study was descriptive statistics which attempted to report the distribution of answer and the difficulty level. The findings of this analysis showed that at least one option of some items was not selected by students, which was likely due to the fact that the option was obviously wrong.

Concerning the reliability analysis, when each component was analyzed separately, the findings revealed that among the three components, listening had the highest reliability coefficient, i.e. α .749 while that of reading and structure and written expression were below .6. However, the reliability of the three components in ETIC was very good, α = .828, which was accepted in this study because ETIC was intended to measure language proficiency that covered all items included in the instrument.

In short, in order to improve the quality of some items, several actions were performed. Firstly we changed some options which were not selected by participants. In addition, for the context of reading

we also looked at the text and modified a little to make sure that the words used did not contain ambiguous meaning. Finally, it should be noted that the items included in the pilot study were exactly the items we used for ETIC and consequently there were no spare items to be removed to improve the quality.

4 CONCLUSION

The main goals of this study were to develop a contextual English proficiency test and apply for its intellectual property rights. To start the process, the steps and principles of test construction were strictly followed by establishing a team who was divided into three following the components included in TOEFL (paper-based).

The first step was the development of test specification, which was then followed by items construction. In the specification, Islam, Indonesia, and academic were decided to be the core themes included in the test and statements instead of questions were applied to start the items included in the test. Before distributing the items in the pilot study, relevant and authoritative experts were asked to review the quality of the items qualitatively, followed by corresponding revision. The next step was a pilot study conducted to students of the university from various majors, which data were analyzed in terms of the difficulty level, the validity and reliability analysis. The findings indicated that the items included in ETIC was valid and reliable to examine English proficiency.

Finally, the process and the findings of this study imply some important suggestions. Firstly, it should be noted that test development is not an easy work. It involves a lot of steps, materials, and times and requires high commitment and coordination. Therefore, it is suggested to prepare a solid team who are committed to the hard work. The team should consist not only test developers but also reviewers and proof readers. Since all the test developers in this study were all Indonesian, it is deemed important to have proof readers to make sure that every single word, phrases and sentences as well as collocations are correct. Secondly, in this study, the item number included in the pilot study was exactly the number of items that would be included in the test. There were no spare items in case deletion was necessary during the analysis. Consequently, there was no item deleted. Although improvement of both the text and the options were conducted, it is considered important to spare some more items to provide some spaces for deletion of

inappropriate items. Hence, it is highly suggested to have more spare items during the pilot study and to keep enriching the test bank.

REFERENCES

- Abdurrosyid, (2017). The Islamic Entries in three major English Dictionaries. *Insaniyat: Journal of Islam and Humanities* Vol. 2, No.1. p. 41 – 49.
- Bachman, L.F., Palmer, A. S. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.
- Brown, H.D. (2004) *Language assessment: principles and classroom practices*. New York: Longman.
- Carr, N. T. (2011). *Designing and Analyzing Language Tests*. Oxford, Oxford University Press.
- Combe, C. & Davidson, P. (2014) Common Educational Proficiency Assessment (CIPA) in English. *Language Testing*, Vol. 31 No. 2, p. 269 – 276.
- Council of Europe. (2011). *Manual for language test development and examining*. ALTE.
- Cohen, A.S. & Wollack (2006) *Handbook on test development*. Madison: University of Wisconsin.
- Chae, E.Y. & Shin, J.A. (2015) A study of a timed clozed test for evaluating L2 proficiency. *English Teaching*, Vol. 70, No 3, p. 117 – 125.
- Chappel, B. (2015) World’s Muslim Population will surpass Christians this century, Pew Says. Retrieved from <https://www.npr.org/sections/thetwo-way/2015/04/02/397042004/muslim-population-will-surpass-christians-this-century-pew-says>
- Chen, Y. & Puttitanun, T. (2015), Intellectual property rights and innovation in developing countries. *Journal of Development Economics*. Vol 78, p. 474 – 493.
- Cheng, L. & Watanabe, Y. (2008) *Washback in language testing*. London: Lawrance Elbaum Associates Publishers.
- Chapelle, C. and Douglas, D. (2006). *Assessing Language through Computer Technology*. Cambridge: Cambridge University Press.
- Creemers, B.P.M. (1994). *The effective classroom*. London. Cassell.
- Creemers, B. & Kyriakides, L. (2008). *The Dynamics of Educational Effectiveness. A contribution to Policy, Practice and Theory in Contemporary schools*. New York & London: Routledge Taylor & Francis Group.
- Cho, H. & Brutt-Griffler, J. (2015). Integrated reading and writing: a case of Korean English language learners. *Reading in a Foreign Language*, Vol. 27, No. 2, p. 242 - 261
- Crystal, D. (2007). *English as a Global Language*. Cambridge: Cambridge University Press.
- Fulcher, G. (2010). *Practical Language Testing*. London: Holder Education.
- Franco, C.P., & Galvis, A. H. (2012). The role of situational context and linguistics context when testing EFL Vocabulary Knowledge in a Language Teacher Education Program: A preliminary Approach. *Colombia Applied Linguistics Journal*. Vol. 15. No 1. P. 85 – 99.
- Hamid, M. O. (2014) World Englishes in Internatinal proficiency tests. *World Englishes*, Vol. 33, No. 2, p. 263 – 277.
- Heaton, J.B. (1990). *Writing English language test*. New York: Longman.
- Hughes, A. (2003) *Language testing for language teachers*. Oxford: Oxford Univeristy Press.
- Kim, E-Y, J. (2018). Utility and bias in a Korean standardized test of English: the case of i-TEPS (Test of Englilsh Proficiency developed by Seoul National University). *Asian Englishes*. DOI: 10.1080/13488678.2018.1463346
- Kopriva, R. J. (2008). *Improving Testing for English Language Learners*. New York: Routledge Taylor & Francis Group.
- Lado, R. (1961) *Language Testing*. London: Longman.
- Marcus, R. (2004) Strategic management of Intelletual property. *MIT Sloan Management Review*, Spring.
- McNamara, T. and Roever, C. (2006). *Language Testing: The Social Dimension*. London: Blackwell.
- Reviere, R., Berkowitz, S., Carter, C. C., Gergusan, C. G. (Eds.) (1996). *Needs Assessment: A Creative and Practical Guide for Social Scientists*. Taylor and Francis: Washington.
- Sabatin, I. (2013). The Effect of cultural background knowledge on learning English language. *International Journal of Science, Culture and Sport*, 1(4): 22-32.
- Soler, E. A., & Jorda, M. P. S. (2007). Introduction. In E. A. Soler, & M. P. S. Jorda (Eds.), *Intercultural Language Use and Language Learning* Springer (pp. 1-6).
- Sosa, K. (2012). Standardized testing and cultural bias. Retrieved from: <http://www.brighthubeducation.com/student-assessment-tools/65699-standardized-testing-and-cultural-bias/>.
- Tomlinson, B. 2010. *Research in Materials Development for Language Teaching*. London: Continuum.
- Top 10 Largets Muslim Populations in the world. (2018). Retrieved from <https://support.muslimpro.com/hc/en-us/articles/115002006087-Top-10-Largest-Muslim-Populations-In-The-World>
- Thaidan, R (2015). Washback in Language Testing. *Education Journal*. Vol 4. No 1, p. 5 – 8.
- Thomas, P & Beriedlid, A. (2015) In the shadow of ‘Anglobalization’ national test in English in Norway and the making of a new English underclass. *Journal of Multicultural Discourse*. Vol. 10, No. 2, p. 349 – 368.
- Phillipson, R. (1993). *Lingustic Imperialism*. Oxford: Oxford Univerisy Press.
- Yan, X. (2014) An examination for rater performance on a local oral English proficiency test: A mixed-methods approach. *Language Testing*, Vol. 31, No 4, p. 501 – 527.,