ELLiC 2019

Proceedings of the 3rd English Language and Literature
 International Conference (ELLiC)

Semarang, Central Java, Indonesia

27 April 2019

EAI

RESEARCH MEETS INNOVATION

Proceedings of the 3rd English Language and Literature International Conference (ELLiC)

27 April 2019, Semarang, Central Java, Indonesia

ELLiC 2019

General Chairs

Dr. Dodi Mulyadi, M.Pd. (Chair)
Robbi Rahim, S.Kom, M.Kom (Publication Chair)
Siti Aimah, M.Pd. (Secretary)
Yunita Nugraheni, M.Hum. (Treasurer)
Muhimatul Ifadah, M.Pd. (Program Chair)
Heri Dwi Santoso, M.Hum. (Editorial)

Conference Organization

Advisor	: Yesika Maya Ocktarani, M.Hum
Chairperson	: Dr. Dodi Mulyadi, M. Pd
Secretary	: Siti Aimah, M.Pd
Treasurer	: Yunita Nugraheni, M.Hum

Divisions

Administration, Secretarat, & Registration	: Riana Eka Budiastuti, M.Pd.
	Kukuh Dharmawan, A.Md.
Event & Entertainment	: Muhimatul Ifadah, M.Pd.
Papers & Proceedings	: Heri Dwi Santoso, M.Hum.
	Budi Tri Santosa, M.A.
Poster	: Testiana Deni W., M.Pd.
Receptionists	: Yulia Mutmainnah, M.Hum
	: Rose Fadilah, S.E.
Public Relations & Sponsorship	: Indah Maftuchah, M.Si. Akt.
Documentation & Publication	: Yunie Armiyati, M.Kep.
	: Anggry W., S.Kom.
Food & Beverages	: Prestiwari Devitri, A.Md.
	Yusvita Nur Andayani, S.E.
Transportation, Accomodation, & Hospitality	: Diana Hardiyanti, M.Hum.
Equipment	: Renny Hersusanti, A.Md.
	Joko Untung
Information Technology	: Mega Pranata, S.Pd.
	Faisal Ridlo Amatullah. S.Kom.
Driver	: Sunar, S.E.

Preface

English Language and Literature International Conference (ELLiC) is an annual event organized by the Faculty of Foreign Language and Culture, Universitas Muhammadiyah Semarang. The event on April 27, 2019 was the third edition of conference, which was held at Unimus' campus on Jl. Kedungmundu Raya No. 18, Semarang.

"Reimagining New Cyber-based Research in English Language, Literature, Linguistics, and Translation" has been chosen at the main theme for the conference, with a focus on the latest research in language teaching, literature, linguistics, and translation.

The conference invited speakers from Thailand, Malaysia, Singapore, and Indonesia, and was attended by more 100 presenters and participants from several countries such as Philippines, Malaysia, Australia, and Indonesia.

Dr. Dodi Mulyadi, M.Pd.

Constructing Test Specification for Standardized Test of Academic English Proficiency (STAcEP)

Rohfin Andria Gestanti[1], Restu Mufanti[2], Elok Putri Nimasari[3]
Universitas Muhammadiyah Ponorogo, Indonesia[1,2,3]

{[1]rohfingestanti@umpo.ac.id, [2]restu.mufanti@umpo.ac.id, [3]nimasari@umpo.ac.id}

Abstract. This present study aims to determine what aspects should be considered in constructing test specification for localized language proficiency test administered in a private university in Indonesia. Employing the CEFR for language as the basic framework, this study involves the analysis of CEFR document and interview with four experts on language testing in order to construct the intended test specification as the basis of a language test before it is further developed. The result of the study implies that there are five aspects with more than ten components in the initial consideration of test specification that should be considered. They include construct of the test, usefulness of the test, content and format of the test, technical features of the test and required standards of the test. The analyzed information regarding those aspects, then, is combined to construct the proposed test specification of STAcEP test.

Keywords: *Test Specification, Language Test, Standardized Test of Academic English Proficiency (STAcEP), CEFR*

1 INTRODUCTION

Test is one of assessment methods to determine someone's ability in understanding and comprehending certain scope of material. Test is purposively designed according to certain objectives, competencies, or outcomes expected by the test provider as well as the test takers[1]. Testing in language learning plays an important role as it convinces to what extend someone comprehend the outcomes of certain materials[2][3]. Rapid development in academic context influences the development of English test as a measurement of one's proficiency. It is proven by the development of localized tests in several countries [4]–[6]. Even though the existing language tests are still considered significant and appropriate to be used, some institutions require the use of certain designated test, particularly English proficiency test. This is due to the different policy that is applicable in different institution[7]. In Indonesian tertiary context, university students are required to have school-leaving certificate that indicates certain competencies in English they have. Such requirement is demanding as many companies urge every job applicants to provide the certificate as the obligation and

requirement in applying the job[8], [9]. As the results, each institution tries to provide English proficiency test for its students. Recently, institutions tend to construct their own test format. Some possible reasons that prompt it include internal institutional policy, consideration on the needs of test-takers and test-users, availability of the test sources and consideration on the expected outcomes. Most importantly, justifying students' needs to the new format of test is possible since it is considered as an effort to analyze their competencies in the most appropriate way based on their characteristics[10]. That is why there is urgency in constructing English proficiency test in accordance to aforementioned reasons.

In the case of English proficiency test in private university where the researchers teach, students are required to enroll in an Academic English Program. According to the institution's policy, all students from all majors should obtain a certificate of Academic English proficiency test as school-leaving certificate[11][12]. This implies that students have to attend proficiency test in order to be certified and graduated from the university. Furthermore, it is also mentioned in the institution's policy that proficiency test administered for students should be able to cover not only general knowledge of English but also local characteristics valued by the institution. In addition, after completing their education, students are supposed to achieve basic level of English proficiency. Simply, certain course and test proficiency program that meets students and institutions' requirement should be constructed[13], [14]. Such proficiency test is yet available in this university. Therefore, there is a demanding concern regarding the construction of English proficiency test. The intended test is called STAcEP or Standardized Test of Academic English Proficiency. This test is initiated to equip students with intermediate level of proficiency. Referring to the institution's applicable policy, the administered language test is supposed to be able to measure students' competency in the level of understanding. Instead of demanding students to produce English in direct communication, this test tends to focus more on receptive skills. The reason underlying this policy is to make the students get use to read, understand, and extract information and references from scientific journals for their final assignment. Therefore, such test is demanding and of high significant to be constructed as an attempt to fill the gap between what test takers and test users need to accomplish. Such situation is similar to previous study in another country which blends localized language assessment with standardized English proficiency [15].

Constructing English proficiency test is not an easy work. It demands rigorous preparation and planning. Decision to construct a test should not be based on mere thought since it involves long processes to match the test with the intended outcomes, characteristics, and standards of standardized proficiency test[16], [17].Despite its significance, having proficiency test that meets the users' needs involves many considerations concerning how the test is formed. Once the decision to provide a test is made, test developers should be informed that they should have some in store regarding the planning of constructing the test. In this case, consulting prevailing, accepted standards of proficiency test development is more than just suggested in order to make sure that the process of constructing the test preparation is aligned and justified with the globally acknowledged, standardized framework[18], [19].

Standardized language test refers to a particular test that has been considered valid as it uses certain framework of test as the basis. In the field of language learning, CEFR for

Language is possibly one of the frameworks that mostly used as a basis framework in developing or constructing test. CEFR or Common European Framework Reference for Language is a framework that suggests a general basis in the construction of language syllabuses, curriculum, courses, examinations, etc that firstly initiated in Europe[20]. As the name suggested, this framework is used as a reference as it depicts descriptors covering what language users can do in a more comprehensive way. Besides, it also provides six-level language proficiency that allows the users to reflect on their competencies. That is why most of the language proficiency tests that available today e.g. TOEFL, IELTS, GEPT, Cambridge ESOL, etc are developed through the application of CEFR[21], [22]. CEFR for language as the reference framework in designing a test has been employed by test developers across countries since past few decades. As some studies suggested, this framework is considered an approach that benefits test developers in constructing the framework of the test as it enables them to adapt the illustrative scales and descriptions of language proficiency level according to the expected outcomes[23]–[25]. According to CEFR, the overall stages in developing standardized test include decision to provide test, developing test specification, producing test, administering test, marking test, grading test, reporting test result, and periodic test review.

As this study is the initial study in the process of developing test, this study focuses on the construction of test specification for the intended test. As what CEFR approach suggested, test specification is the basic phase that provide detail information of the test including the purpose, significance, format and content, technical features, and requires standard of the proposed test[26]. In this phase, test developers are supposed to be well informed concerning the detail components of the test. These details are commonly addressed with initials considerations[16]. Thus, several points are addressed so that the intended test specification can be constructed. The points include the focus of the tasks, skills or competencies intended to be tested, text types as the input of the test, text sources, and areas of topic considered suitable for the test and intended level of competency. Moreover, some other points regarding the test' technical features are also addressed. It includes the duration, number of items, types of items, length, format and marking system of the test [16][20]. Detail information regarding the components derived from the manual for language test development and examining is the main material to be analyzed in order to construct the test specification. The aforementioned characteristics are based on the CEFR approach as the mainly underpinning theory. In addition, the underpinning framework is also the tools by which the requirements of localized language assessment of the current institution are aligned to. Test specification is an urgency that should be constructed first before the actual test is developed[24]. Based on the given explanation, this study is primarily conducted as an attempt to answer a concern about initial considerations in constructing test specification for STAcEP that is based on the indicators of CEFR.

2 METHODOLOGY

In carrying this study, descriptive qualitative approach by elaborating the designing process of test specification suggested by CEFR for language was employed[16]. According to this framework, designing test specification is carried out by determining the test contract, test

usefulness, content and format, technical features and required standards of the test. The next step is consulting CEFR descriptors to obtain the judgment between test specification and the CEFR level. Those overall steps were conveyed in the form of description and elaboration. In this study, four experts of language testing with reputable credentials in the field of language testing are involved as the participants. In this case, insights from the experts were concluded as the judgment for STAcEP test specification. In collecting the data, the researchers employed document-based observation and focus group discussion with experts. The main data in this study is CEFR document. Document-based observation is conducted by participants and researchers within a particular discussion[27]. The observation results then are crosschecked through focus group discussion with the experts to compose judgment for the test specification. The result of the analysis is described in tables in accordance with the parts of initial considerations of constructing test specification. Thus, it is interpreted and reported in narrative and then concluded as the result of the study.

3 FINDING AND DISCUSSION

The results of this study are presented in tables representing the initial considerations of intended test specification suggested by CEFR. There are three parts of initial considerations reported in this study: (1) construct, usefulness, and content and format of the test; (2) technical features of the test; and (3) required standards of the test.

Table 1 Result of Initial considerations of STAcEP test specification in regards to the construct, usefulness and content and format of the test

Aspects	Components	Document-based observation	Discussion of justification
Test construct	Skill to be tested	**Semantic components** a. lexical semantic: related to word in general context and inter-lexical relations b. grammatical semantic: the meaning of grammatical elements c. pragmatic semantic: entailment, presupposition and implication	Receptive skills are more suggested for users in non-English speaking country for the objectivity of the assessment
Test usefulness	Focus of the test	Providing basic level of English proficiency for university students	Test that allows students to obtain simple information of general English
Content and format	Focus of the tasks	**Listening tasks:** a. Retrieving factual, specific information b. Following instructions and directions c. Recognizing numbers or letters **Reading tasks:** a. Identifying topic of the test b. Retrieving specific information by scanning the text c. Locating and selecting relevant information to complete the task d. Making inferences from the	**Listening tasks:** a. Tasks that allow test takers to pay attention to every piece of given information in the text b. Tasks that require students to comprehend the particular part of the text **Reading tasks:** a. Tasks that allow students identify the purpose and general meaning of the text b. Tasks that allow test takers to pay attention to every piece of given information in the

		information given in the text	text
			c. Tasks that require students to comprehend the particular part of the text
	Text types and sources used as input for the tasks	**Input for listening:** a. Public lectures b. Public announcement c. Inter-personal conversation **Input for reading:** a. Newspaper and magazine articles b. Non-fictional books c. Research reports d. Journal articles e. Paper assignments	**Input for listening:** a. Cross-gender conversations between two or more people in real-life setting allow test takers grasp information in the closest way b. Authentic materials from public services or announcements enable students to understand text in real situation. c. Monologues that is similar to lectures allow students analyze formal-informal situation **Input for reading:** a. Texts taken from magazine or newspaper are more familiar for students and have various ranges of lexical choices b. Texts taken from scientific journals (research reports or assignment) enhance students' understanding of particular context with more formal word choices c. Texts taken from credential websites may be useful to develop comprehension of various topics unavailable in books or articles
	Topic areas considered suitable for use	a. Basic needs in daily life including services, shopping, places, places and direction, etc. b. More formal topics including education and job interview c. Social related topics such as relation with other people	a. Topics related to basic information of daily life covers how to obtain and find information (daily-activity-based topics) b. Texts covering hard discussion i.e. work field, education etc. may widen students' knowledge about formal text c. Muhammadiyah and Islamic values depicts the institution' Islamic character d. Topics related to local cultures or values to prove that the institution highly value the local characteristics in a globalized context
	Format of the tests	Discrete-item test or task-based test items depends on the intended focus of the overall test	Test items are supposed to be based on several text types with particular numerous instructions, items and responses

According to table 1, there are six components that should be considered by test developers in constructing the STAcEP test specification. The first component in regarding what skill or competency is intended to be tested. The observation result reveals that semantic components are considered suitable competencies to be tested in STAcEP. Semantic components are components of competencies that refer to briefly understand spoken or written text, analyzing word meaning, etc[20]. Such competencies are led to the involvement of receptive language skills. This consideration is determined by consulting the general aims of the intended test and the institutional policy regarding language test. According to the policy,

students are expected to understand English communication whether for spoken or written texts. Therefore, semantic components including lexical semantic, grammatical semantic and pragmatic semantic are appropriate competencies for the test. Moreover, experts' justification revealed that test focusing more on the using of receptive skills of English is considered suitable. The current institution is not using English as instructional language, thus, testing oral or written productive language skills will be hard in term of the assessment. This statement is in line with the result of studies analyzing assessment for receptive and productive language[28], [29]. Referring to this result, it can be inferred that competencies fall under receptive language namely listening and reading are the skills to be tested in STAcEP.

The second component is related to the usefulness or general focus of the test. This aspect is closely related to the expected outcomes of the test itself. Regarding this component, it revealed that STAcEP is purposively designed to be used for university students in a private institution in Indonesia. Therefore, providing basic level of English proficiency is adequate, particularly to obtain school-leaving certificate. Experts conveyed that a localized yet standardized language test should ponder the test takers and their current level of proficiency. Consulting the factual situation, providing a language test that allows students to analyze texts to obtain simple, particular information is significant. Therefore, STAcEP is designed with expectation that its test takers could achieve at least basic level of proficiency. In CEFR itself, basic level of proficiency means that the test takers are able to understand basic information on daily basis[28].

The next component is the focus of the tasks of the skills chosen to be tested. As what is discussed before, listening and reading are two main competencies of STAcEP. Therefore, the focus of the tasks is defined. Generally, the focus of both skills is quite similar since both tasks are inclined to provide relatively related outcomes. As table 1 showed, listening and reading tasks generally focus on obtaining particular information from the texts. However, there are some focuses that are different, e.g. the instructions, the use of numbers and letters for listening tasks and topic identification, information selection and inferences making for reading tasks. Such differences are possible, even are supposed to, since the students' expected competencies after completing the tasks are also different. Each task is purposively chosen in order to make students achieve certain goals. These results are also supported by the experts' judgments in which there should be various tasks in the test with different purpose. By doing so, it may enable students to enhance their knowledge and achieve the expected competencies as they are able to finish those tasks. In this case, the focus of the tasks is decided by considering the general outcomes or competencies required by the institution that develops the test.

The fourth component is the text types used as the input for the tasks. Table 1 displayed that there are several types of texts that are supposed to use in the construct of STAcEP test specification. By consulting the document of CEFR and reflecting into institution's requirements, some types of texts can be taken from various i.e. public lectures, public announcements, and conversations for listening tasks and articles taken from magazine, newspaper, scientific journals, research papers, etc. Those types of texts are included in STAcEP test since they cover various texts that allow students improve their knowledge.

Expected competencies are the basic justification to select such sources. The same result also goes to the justification from the experts. For listening text types, the experts assert that input taken from conversation with several turn-taking role allow test takers to grasp the information in the closest way and input from monologue or lectures allow students to analyze the use of language depend on the formality of the situation. For reading tasks, participants generally have similar results with the CEFR suggestion. However, another type of text may be considered to be used; texts from credential websites. They are beneficial to use since they offer wide range of texts that books or other sources do not cover. Using various text types from different sources may enhance the text validity as it helps students understand texts in many areas [30].

Component showing the topic areas considered suitable for the test is the next component analyzed in the study. Basically, theory suggests that topics related to daily basis or necessities and social related topics are significant to be involved. It includes goods and services, places and directions, shopping, work life, job interview, relation with other people, etc. Besides, topics that are potential for formal discussion such as education or religion are also of high significant. They provide content, context, lexical choices, analysis and scope of discussion that lead students to a broad comprehension of intended competencies[31]. This finding is supported by the justification given by the experts who convey that topics related to daily basis, social or education will stimulate students' understanding as the topics are some things they can relate to in their daily life. In this case, some experts assert that localized language test sometimes make use the involvement of local-related topics. As it has the same intention with the requirements set by this university, therefore, topics related to Islamic values and local culture are included the test. Not only due to the fact that this university is an Islamic university and located in small town in Indonesia, but introducing local characteristics in English language test may build students' understanding and broaden their knowledge on the role of local issues in global setting and vice versa.

The last component in the first part of initial considerations of STAcEP test specification is format of the test. It refers to the overall types of test items included in the test. Table 1 revealed that using discrete-item test or task-based-item test is appropriate. However, this result is not considered applicable for STAcEP since using more than one format for a test is not in line with its requirements. In addition, it may not suitable with the intended competencies that are expected by the university. According to the justification, experts suggest that using one format and focusing on that is more significant. Referring to the objectives of the test, employing test format based on several types of texts with particular instructions, items and responses is more acceptable and in line with the outcomes. Therefore, task-based-item test tend to be used in STAcEP test specification.

The second part analyzed in this study is the technical features of the test that is displayed in table 2

Table. 2 Result of Initial considerations of STAcEP test specification in regards to technical features of the test

Aspects	Components	Document-based observation	Discussion of justification
Technical features	Duration of the test	Depends on the consideration of test takers' ability in completing the test	Approximately no less than 2 hours, depends on the numbers of test items and types of intended response
	Estimated number of items in test	Depend on the test length (number of sections, length of text per section and number of item per section)	Depend on the test length (number of sections and number of item per section)
	Estimated number of items per section	50 – 80 items	50 – 80 items
	Types of item	Selected response (multiple choice, matching, ordering) Constructed response (short answer/ gap filling, extended writing)	Various types of items let the students elicit intended response from the provided input and tasks' focus. The items may include multiple choices, true-false answer, matching, completion, short answer, etc.
	Length of text	Measured in words or minutes for input in spoken text and measured in words for input in written text	Measured in words for oral and written input to check the relevance of the script and intended items and responses
	Marking types	One mark per item or weights the items based on the focus of the task	One mark per item since the test is intended for receptive skills

Table 2 showed that six components fall under the technical features of the test. The first component is the duration of the test. After analyzing the CEFR document, it is revealed that the appropriate duration of the test may vary depending on the ability of test takers in accomplishing the tasks. It means that several try outs of the drafted test specification should be conducted in order to get the exact appropriate duration. However, according to the experts' justification, the result showed that the estimated duration is approximately not less than two hours. Such justification is based on the numbers of test items and types of intended response. Enough time limits are supposed to be allocated for average test takers to finish the tasks within the mean time. However, since this STAcEP test specification is not piloted yet, the consideration of duration of the test is relied on the experts' justification.

The next two components are regarding the estimated number of items for the overall test and per section. As what is displayed in table 2, it can be seen that number of items for the whole test are various depending on the number of sections, length of text per section and number of item per section, while the suggested number of items per section is 50-80 items. According to this result, defining the number of test items per section should be conducted first before defining the number of test items as a whole. Considering it, the experts gave quite exactly the same justification. It can be inferred that STAcEP test specification is constructed by considering the suggested result.

The fourth component of STAcEP test specification is determination of test items that will be used. It is probably one of the most important parts in constructing language test and it is crucial in construction process. From the result illustrated in table 2, it can be inferred that the selection of the types of test item is closely related to the type of expected responses to be

given since different responses carry different purposes the students should be able to accomplish. Based on the displayed result, types of test items that are used are the combination of selected response and constructed response. Moreover, the use of various types of items let the students to elicit intended response from the provided input and intended focus of the tasks. Simply, the types of items are ranging from multiple choices, true or false answer, matching, and completion to short answer.

Length of the text used in the test is the next component analyzed in initial considerations of test specification. In this component, the text used in the test should be measured in term of its length in order to specify its content and maintain its reliability. As what CEFR indicated, the length of the text is measured in minutes by determining particular length of time when the text is orally recorded or in words by setting certain average number of word count for the written text. However, according to the experts, it can be seen that using the same technique to measure the length of the text is more suggested. This is done in order to manage its consistency. Thus, the texts used in STAcEP test specification are going to be constructed by determining the word count.

Marking types employed for the test is the last component in initial considerations of STAcEP test specification. It may be the last component for this aspect, but it is one of the most crucial parts. Marking is the activity of giving score for the responses given by the test takers. In this study, using one mark per item is more considered as the objective of the test itself is to test receptive skills. Generally speaking, weighting the items is also available for marking system. However, it may involve rating scale as well. Since the present study focuses more on testing receptive skills and deliberately attempts to manage the validity of the score, objective marking system by giving one mark per item is employed.

The last part of construction of STAcEP test specification is the aspect of institutional required standards. Table 3 depicts the result of the analysis for this aspect.

Table.3 Result of Initial considerations of STAcEP test specification in regards to required standards of the test

Aspects	Components	Document-based observation	Discussion of justification
Required standards of the test	Level of performances	Six levels of competencies (A1, A2, B1, B2, C1, C2) It depends on the objective of the test and expected outcomes	Basic level or pre-intermediate level may be appropriate to be applied in the institution as it considers its objectives and applicable policy
	Characteristics of rating scale	Relative of absolute scales as it is adopted from CEFR levels and scales	CEFR descriptors for proficiency level are provided under particular terms along with the description for each term.

Table 3 shows the analysis result of initial considerations in constructing STAcEP test specification in term required standards of the test. The first component is regarding the level of performances expected to be achieved by the students. CEFR is a framework to construct language assessment that provides six levels of language competencies along with the descriptors. According to this framework, constructing level of competencies should be based on these levels by taking into consideration the objective of the test and expected outcomes. Consulting the policy and the requirement of the university, STAcEP is aimed to provide

students with the basic language competency so that the students may have basic level of proficiency. Simply, this test is in the scope of understanding level. In the CEFR descriptor, basic level of understanding is placed in A2-B1 descriptors. By considering the descriptor is this level, level of competencies for STAcEP is constructed.

The second as well as the last component is defining the characteristics of rating scale. Rating scale here refers to the approach to compose the descriptor for the levels. As what is shown in table 3, relative of absolute scales s the rating scales that is used in constructing test specification for STAcEP. Relative or absolute scales are appropriate in this case since this scale is CEFR-adapted scale that is described in words to define performances in definite terms [16]. It also has been discussed in previous part that CEFR levels are provided with its descriptors. Such scales and descriptors are illustrative and adaptable to be used in different context and setting. Moreover, it is not definitive so the levels of language competency may be ranging based on the result of the test. Therefore, in order to make STAcEP test be a standardized test, the intended outcomes of the text derived from the institutional policy and characteristic is described by adopting CEFR descriptors in the same level.

4 CONCLUSION

Decision of constructing language test is initiated by the consideration of the test specification as the first basis in developing the test. By employing CEFR for language as the basis framework, then test specification for STAcEP is constructed. There are three aspects taken from the framework covering the general form of test specification. The result showed that there are about 14 components revealed from the documentation and experts' justification that compose STAcEP test specification. However, there is a limitation of this study regarding the content validity and reliability. Thus, further research focusing on piloting and analyzing this concern should be conducted in order to address the validity of the test as well.

Acknowledgments

References

[1] C.-M. Shih, "How tests change teaching: A model for reference," *English Teach. Pract. Crit.*, vol. 8, no. 2, pp. 188–206, 2009.

[2] R. Mufanti, A. Susilo, R. A. Gestanti, and E. P. Nimasari, "THE EFFECT OF CONSTRUCTING AND ANALYZING MODEL IN ENHANCING TERTIARY EFL LEARNERS ' GRAMMAR," *PEOPLE Int. J. Soc. Sci.*, vol. 4, no. 3, pp. 1896–1911,

2019.

[3] B. Y. J. M. Norris, "Purposeful Language Assessment : Selecting the Right Alternative Test," *English Teach. Forum*, vol. 38, no. 1, pp. 41–45, 2012.

[4] N. Rahmat, L. S. Min, N. A. M. Sungif, and F. N. M. Yusup, "English Language Proficiency Tests and Academic Achievement : A Study on the Malaysian University English Test as a Predictor of Technical Programme Undergraduates Academic Achievement," *Adv. Lang. Lit. Stud.*, vol. 6, no. 1, pp. 114–119, 2015.

[5] M. Wu, "Comparing PETS and GEPT in China and Taiwan," *English Lang. Teach.*, vol. 5, no. 6, pp. 48–53, 2012.

[6] Y. Tsai and C. H. Tsou, "A standardised English Language Proficiency test as the graduation benchmark : student perspectives on its application in higher education," *Assess. Educ. Princ. Policy Pract.*, vol. 16, no. 3, pp. 319–330, 2009.

[7] E. Shohamy, "Language tests as language policy tools," *Assess. Educ.*, vol. 14, no. 1, pp. 117–130, 2007.

[8] Rahmi, "THE DEVELOPMENT OF LANGUAGE POLICY IN INDONESIA," *Englisia*, vol. 3, no. 1, pp. 9–22, 2015.

[9] S. S. Mappiasse, A. Johari, and B. Sihes, "Evaluation of English as a Foreign Language and Its Curriculum in Indonesia : A Review," *English Lang. Teach.*, vol. 7, no. 10, pp. 113–122, 2014.

[10] Y. Jin and L. Hamp-lyons, "A new test for China ? Stages in the development of an assessment for professional purposes," *Assess. Educ. Princ. Policy Pract.*, vol. 22, no. 4, pp. 397–426, 2015.

[11] R. Mufanti *et al.*, "PUBLIC SPEAKING INTENSIVE COURSE FOR STUDENTS OF NON- ENGLISH DEPARTMENT (A Study in English Language Centre)," *Expo. J.*, vol. 7, no. 2, pp. 161–178, 2018.

[12] U. M. Ponorogo, "Renstra Universitas 2017," Ponorogo, 2017.

[13] E. P. Nimasari, "An ESP Needs Analysis: Addressing the Needs of English For Informatics Engineering," *J. Educ. Soc.*, vol. 1, no. 2, pp. 1–16, 2016.

[14] R. A. Gestanti, E. P. Nimasari, and R. Mufanti, "ESP ISSUE IN INDONESIAN TERTIARY CONTEXT : WHAT STUDENTS NEED IN LEARNING ENGLISH," *PUPIL Int. J. Teaching, Educ. Learn.*, vol. 3, no. 1, pp. 98–117, 2019.

[15] L. Liu and G. Jia, "Looking beyond scores : validating a CEFR- based university speaking assessment in Mainland China," 2017.

[16] C. of Europe, *Manual for Language Test Development and Examining*. Strasbourg: Association of Language Testers in Europe, 2011.

[17] GoFLUENT International Inc., "A Guide to English Tests and Proficiency Frameworks," New York, 2012.

[18] J. M. Sims, "A Valid and Reliable English Proficiency Exam : A Model from a University Language Program in Taiwan," *English as a Glob. Lang. Educ. J.*, vol. 1, no. 2, pp. 91–125, 2015.

[19] C. Shackleton, "Developing CEFR-related language proficiency tests : A focus on the role of piloting," *Gruyter Mout.*, vol. 8, no. 2, pp. 333–352, 2018.

[20] C. of Europe, *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Stasbourg: Cambridge University Press, 2001.

[21] U. of C. E. Examinations, "Using the CEFR : Principles of Good Practice," 2011.

[22] J. Dunlea, "Incorporating the CEFR into language test development: Using an international framework in local contexts," Moscow, 2014.

[23] S. Arnott, E. Piccardo, K. Rehner, and S. K. Taylor, "The Common European

Framework of Reference (CEFR) in Canada : A Research Agenda," *Can. J. Appl. Linguist.*, vol. 1, no. 20, pp. 31–54, 2017.

[24] L. P. Division, *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR) A Manual*, no. January. Strasbourg, 2009.

[25] M. Wernicke and M. Bournot-trites, "Introducing the CEFR in BC : Questions and Challenges," *Can. J. Appl. Linguist.*, vol. 2, no. 14, pp. 106–128, 2011.

[26] ALTE, *Materials for the Guidance of Test Item Writers*, no. July. ALTE, 2005.

[27] T. Grammes and M. Açıkalın, "Qualitative Data Collection and Interpretation : A Turkish Social Studies Lesson," *J. Soc. Stud. Educ. Res.*, vol. 7, no. 1, pp. 1–40, 2016.

[28] S. Bailly, S. Devitt, M. Gremmo, and F. Heyworth, "Common European Framework of Reference for Languages : Learning , teaching , assessment."

[29] D. Renukadevi, "The Role of Listening in Language Acquisition ; the Challenges & Strategies in Teaching Listening," vol. 4, no. 1, pp. 59–63, 2014.

[30] E. Haliza, E. Ibrahim, I. Sarudin, and A. J. Muhamad, "The Relationship between Vocabulary Size and Reading Comprehension of ESL Learners," *English Lang. Teach.*, vol. 9, no. 2, pp. 116–123, 2016.

[31] E. P. Nimasari, "CRITICAL DISCOURSE ANALYSIS IN TEACHING READING Elok Putri Nimasari," *J. English as a Foreign Lang.*, vol. 6, no. 2, pp. 115–126, 2016.