How foreign speakers implement their strategies to listen Indonesian language?

Memet Sudaryanto[1], Doctoral Student of Post Graduate Program in UniversitasNegeri Yogyakarta, Indonesia

DjemariMardapi, Evaluation and Research Education of UniversitasNegeri Yogyakarta, Indonesia

SamsulHadi, Evaluation and Research Education of UniversitasNegeri Yogyakarta, Indonesia

Abstract- This study aims to describe the ability of foreign speakers, especially the listening skill/ability. To achieve this objective, we need an appropriate instrument to describe the test takers' abilities well. The employed development model was a modification of the Borg and Gall's and the development of standardized test. The validity was tested through a confirmatory factor analysis (CFA) with P-value = 0.185 and the Aiken's validity of each item met the Aiken's table threshold. The reliability of the test was estimated by Cronbach's alpha with alpha value= 0.70, while the construct reliability, omega, and maximum utility were fulfilled of more than 0.7.The study fulfilled the assumption of the IRT; unidimensionality, local independence, and parameter of invarince. Using Rasch Model to analyze the test items, this study resulted in the difficulty indices of listening test items ranging from -2.29 to 3.64. The results of the competency assessment indicated that 28% of the learners are at basic level, 39% at intermediate level, and 33% at advanced level. The percentage of test-takers' ability (foreign speakers) indicates that the test-takers possess low ability in responding to the rules (structure).

Keywords: Foreign Speaker, Listening, Competency

Introduction

One of the language internationalization tools is a standardized instrument used to measure test takers' abilities. Foreign speakers who want to stay in certain countries are required to be proficient in using the target language. It assumed that receptive skills the capital to develop productive skills (speaking and writing skills). Kim and Gilman state that someone who has weak listening skill naturally has difficulty in speaking, for example difficulties in pronunciation, rules, and understanding of information(Kim & Gilman, 2008). This is supported by Elder who explains that the receptive and productive skills package must be developed partially. After the receptive skills are mapped accurately, the productive skills instruments can be arranged with the same objective(Elder & McNamara, 2015).

Language assessment standards are not only about the preparation of good items, but also, more specifically, about the determination of standard settings. Determination of competency/performance category limitations is one of the most important things in the development of tests, administration, and assessment reports. The assessment climate in Indonesia does not seem to be fully prepared with the application of scientific measurement. Aiken describes that the determination of graduation standards is still far from public understanding(Aiken, 1985).

Some important points regarding the method of determining standards or competencies are still not understood, such as (1) assessment of items through a focus group discussion (FGD), (2) different methods of determining the tiered graduation standard with a cut score, and (3) the position of experts as one of the means of determining the graduation standard. Saddhonoadded that judgment on the conceptual framework is aimed at checking the quality of the assessment after the threshold has been obtained(Saddhono, 2015). Currently, the threshold of the BIPA (Indonesian Language for Foreign Speaker) students' graduation competency is only determined based on the agreement of the decision makers (stakeholders). This is as what is stated by Khairani and Razakthat the determination of standard settings in schools often uses the threshold to evaluate students based on their level of performance(Khairani & Razak, 2013).

The process of determining graduation standards is important to be completed with (1) strengthening the construct to be tested, (2) writing valid and accountable item questions, and (3) a continuous evaluation process to obtain the most valid standard. Indonesian language indeed has aspects of assessment that are different from other theoretical subjects. The standard pattern of Indonesian in everyday use is for learning and service purposes. Indonesian language for academic purposes has a different character from the purpose of communication in

general. Therefore, the important aspect of assessment that becomes the focus is the use of Indonesian in a formal variety.

The proportion of the standard settings of each language skill also has a different position. From the skills tested, the assessment is not merely done by making a balanced proportion between scores obtained from each language skill because the skills tested have different characteristics. Saddhono and Slametexplain the characteristics of each language skill is a partial competency that supports each other(Winarni, Slamet, & Saddhono, 2018). The skill to respond to rules is a critical-analytical ability, not a receptive/productive skill anymore. Listening and reading skills as receptive (receiving responses) skills have different characteristics, especially, from other types of productive skills such as speaking and writing. The estimated score for each skill needs to be standardized in standard rules using the multilevel standard method.

Composite scores of Indonesian language test results are sometimes not applied ideally according to the proportion of their respective needs. The results of the measurement of each language skill and the study of the rules of responses are a form of interpretation of one's language skills. One of the validation techniques used is through composite scores to determine data validity based on different indicators. The main problems in this study are (1) the design and characteristics of Indonesian language competency test instruments that are valid and reliable according to the competencies of foreign students and (2) no proportionality of composite scores represented in the standard setting that is appropriate for foreign speakers. The statements certainly support the main idea that the mastery of Indonesian requires a standard reference that can represent one's language competence.

Language is the only communication medium that humans use to give and receive responses. As a media, language is divided into two, namely written language and spoken language. As a form of skills, language is divided into productive and receptive skills. Terrel and Brown states that listening, and reading are receptive. Listening and reading skills are in the category of receptive skills while writing and speaking skills are productive skills(Terrell & Brown, 2006). This classification underlies the function of language as a medium of communication.

In addition to the communication function, language has a central function as the cultural identity of the people. Indonesia recognizes Indonesian language as the national language, the language of unity, and the language of communication. These three functions underlie the importance of nationally and internationally standardized and applicable Indonesian language. Moreover, currently Indonesian is being pioneered as an international language in the Asian Region.

The main points of language identification are divided into four skills namely, listening, speaking, reading, and writing skills. Listening skills are closely related to speaking skills as reading skills are integrated with writing skills. This is in line with the statement of Cunningsworth that the basic knowledge of language includes grammar, lexemes, and four language skills such as listening, speaking, reading, and writing, all of which must be learned simultaneously(Cunningsworth, 1995).

Language has integrative characteristics, in which each skill cannot stand alone. Language users must train their receptive skills to improve their productive skills. Weak receptive skills result in weak productive skills and vice versa. The result of a good learning process can be seen from the productive skill proficiency. Writing skills as the peak of language competence are ultimately considered the most complex and most difficult skills. Tarigan (2008: 2) states that inter-language skills have a close relationship. The learning process of one skill is training three other skills. The main point of the skills being tested is the receptive skill, which is based on the rational importance of mastering receptive skills to maximize the development of productive skills. The skills studied include listening skills, reading skills, and skills of giving response to rules.

Brown explains that the product of listening and reading skills is not speech or writing, but the result of structured listening assessments in the brain and the task of test takers to observe (assess) the results of the absorption of test takers' information(Terrell & Brown, 2006). Improving listening skills also means helping improve the quality of one's speech. A good listening concept is listening comprehension because it requires the skills to memorize, use, and master several facts they listen to, so listening helps to capture the main idea proposed by the speaker.

The focus of listening skills being tested is the variety of intensive listening, which is divided into several types namely, critical listening, concentrated listening, creative listening, exploratory listening, interrogative listening, and selective listening. Performance on listening skills is invisible and inaudible, but the process of absorbing meaning from what one listens is transferred through the ear and brain. Listening skills require assessment instruments that have reasonability to make the leaps needed through the process from conclusions, from the acceptance of un-observed conclusions about comprehension competencies (Saddhono & Rohmadi, 2014). Listening skills are ultimately the basis for understanding one's language in direct speech.

356

Some basic forms of listening skills according to Brown are classified into four forms, (1) recognizing the form of sound and storing it for a while in short-term memory, (2) simultaneously determining the form of speech (monologue, dialog interpretation, and transactional dialog) which is processed and adapted to the context (speaker, location, and purpose), as well as the content/message of the speech, (3) using top-down techniques to describe the linguistic meaning to interpret the explicit and implicit messages of the speech, and (4) in many cases, omitting linguistic forms from the message received for only the relevant concepts/information(Brown, 2012).

Research Method

This study is a design and development study. It is focused on making an assessment instrument of listening, reading, and responding to the rules assembled as an Indonesian language competency test instrument to measure the language skills of foreign speakers. Testing was done after the Indonesian language learners had followed learning (both formal and informal curriculum). After the instrument had been tested on foreign students, a focus group discussion was conducted to determine the standard setting with three levels of student achievements(Cizek & Bunch, 2007).

The preliminary study was conducted by (1) analyzing the needs of the Indonesian language competency tests with a qualitative approach, namely interviews and observations at BIPA organizing colleges to see the importance of developing an Indonesian language competency test instrument as a standardized competency test in Indonesia for foreign students, (2) analyzing the Indonesian language competency used by several universities, namely collecting, reducing, and concluding materials with the needs of test materials on the instrument produced, and (3) designing test materials with qualitative confirmation in the form of a focus group discussion with a team of experts who determine what material to be tested based on the book involving foreign speakers teachers and researchers. The design of the test material was carried out with snowball responses, a method of gathering information in the form of test materials by continuously asking for experts' opinions until reaching the saturation point of the data(Hambleton & Swaminathan, 1985).

The Listening Section is used to measure the ability to understand information expressed verbally, both in the form of dialog and monolog. The types of dialogs and monologs chosen in the instrument correspond to the needs of students to continue their studies. The conformity is seen from the type of utterance, utterance function, and utterance usage. This section consists of 29 multiple-choice questions with the time allocation of 25 minutes.

Reliability is the consistency of the instrument in measuring the language competence of foreign students after participating in the Indonesian for foreign speakers program.The estimated reliability of the instrument was determined by calculating the Cronbach's Alpha correlation coefficient for the test results(Retnawati, 2008). The internal consistency of the test was carried out only once on a group of subjects, followed by the analysis or estimation of the reliability value. In general, we can use the Alpha Coefficient formula to estimate this reliability. The conformity of the item characteristics with the type and purpose of the test largely determines the quality of the test item.

Result And Discussion

The Standard Setting referred to in this study is a multilevel standard, which determines a threshold by making a range. The threshold in this study does not aim to determine the passing of test takers, but rather places the test takers on certain criteria. Thus, it is recommended that the next study determine the treatment to test takers at different levels. Standard setting procedure with multilevel began with the data prepared in advance. The questions had been arranged from the most difficult to the easiest because by doing so panelists found it easier to determine and enter to what page the students are considered capable of answering questions correctly as the standard setting determination. Khairani, Razak(2014: 198) state that the bookmark method uses the item mapping procedure, where all items are presented in the form of rank-order using a booklet (Ordered Item Booklet). This study examines the determination of Indonesian language competency test setting standards with a multilevel standard approach or tiered graduation standards to determine the level of ability of Indonesian language as a foreign language learner(Cizek & Bunch, 2007).

This study focuses on measuring the ability of Indonesian language use, especially receptive skills for foreign speakers. The specifications of the study of the scope of Indonesian language materials are (1) second language acquisition, (2) receptive skills and skills of giving response to rules, while the specifications in the field of measurement are seen from (3) the development of language test instruments, (4) re-determination of Indonesian language competence in receptive skills , (5) measurement of item difficulty level, and (6) description of language competence in foreign speakers. Based on the specifications constructed, the variables of the study are divided into three dimensions, each in the form of an Indonesian language test instrument for foreign speakers, a

*Corresponding Author: Memet Sudaryanto， Email: memet.sudaryanto2016@student.uny.ac.id

panelist who is a teacher who has a teaching period qualification, and a description of the characteristics of test takers.

Each competency designed in both books is written in a checklist based on the urgency/importance of the competency being tested. In addition to the test material available in the BIPA textbook, the instrument is equipped with an additional test material column used to complete the important test material and is not included in the BIPA textbook.

In the listening domain, the test is in the form of a skill of receiving daily communication responses. The expected learning experiences from this domain are (1) receiving responses, (2) understanding responses, and (3) giving responses coherent with responses. Listening with academic purposes is specified from the simple to complex recordings. The form of recordings used are monologs and dialogs with various themes according to academic purposes. The basic listening domain is in the form of responses to utterance without context while the complex listening domain is at the level of discourse understanding and interpretation.

The test indicator framework was tested through panelists who ate BIPA teachers with 5-year experiences of teaching Indonesian language for foreign speakers.

Table 1. Cycle of Indicator Construction

Cycle 1	Panelists gave an urgency rate to each test indicator item to check how important the test item to be tested to foreign speakers.
Follow-Up	The researcher reduced and added test indicators after being tested to the panelists. The indicator would be omitted if the panelists' average urgency rates is below 3. The panelists can provide additional test indicators through the in-depth interview process and panelist triangulation.
Cycle 2	The panelists would give urgency rate to indicator items after reconstruction based on the average value of items. The indicator I framework was revised based on the panelists' inputs, i.e. adding essential indicators and reducing non-essential indicators.
Follow-Up	The indicator reconstruction is based on the panelists' inputs by collaborating each panelist opinion separately to regulate independence.
Cycle 3	The indicator finalization with an assessment of the importance of items was followed by in depth interviews to re-check the results of reconstruction of cycles 1 and 2.

After the indicator construction, the theory triangulation was conducted to arrange the outline that forms the test indicators. The reconstruction process of the test item indicators ended with a focus group discussion with seven foreign speakers who are in Indonesia for working, education, and cultural exchange purposes. The FGD discussed the users' perspectives regarding the use/purpose of test instrument for foreign speakers.

The integrated understanding in the reading and listening processes is more focused on phrases, sentences, or paragraphs rather than on a single word. Listening is the process of capturing verbal symbols with full attention, understanding, appreciation, and interpretation to obtain information, capture content, and understand the meaning of communication through speech. The success of the listening process is the accuracy of interpreting the recording obtained by the test takers.

Listening as a process of transforming language sounds requires responses in the forms of identification, interpretation, and reaction of stimuli. One of the difficulties presented by the panelists regarding listening skills is the diverse responses that require foreign speakers to interpret based on the context they accept. Listening is the process of receiving information through speech by the sense of hearing carefully. Some variables of the listening materials that need to be considered are (1) the number of narrators, the more narrators who contribute to speech, the simulations will be more complex, (2) the topic of the recording may not contain discriminants that burden or facilitate certain test takers, (3) the clarity of the listening materials divided into clarity of articulation, pauses, intonation, and supporting tone of speech including the voice background being played. In addition to the recording variable that needs to be considered, the listening test developer should emphasize the concept of listening as a process of studying the information received to carry out activities on the content of the recording that has been interpreted (in this context are foreign speakers).

In general, the instrument was developed through two processes, namely the creation of a written test framework and an audio test as a stimulus. The audio test is in the form of dialog and monolog, each of which is made randomly based on the developed test indicators. The test package consists of test audio, question sheets, answer choices, and answer sheets. The listening instrument consisted of 29 test items developed from the

358

question indicators. The test takers worked on the instrument within 20 minutes. Each item has been adjusted according to the working time of each item, which is in accordance with the needs of foreign speakers to respond to the answer by keeping in mind the length of the question and the length of the answer choices.

Table 2. Listening Test Indicators

Competence	Test Indicators
Listening to simple sentences	Showing parts of the dialog and monolog about greeting, farewell, thanking, and apologizing.
Listening to descriptive speech	Showing parts of the monolog and dialog about self-introduction, conversations of family members, mentioning days/dates/months, characteristics/traits, and room description, information in news and public service advertisement.
Listening to the exposition speech	Identifying information in the form of occupations, public transportation, tourism, national figures, commercial advertisements, reports and commands.
Listening to argumentative speech	Identifying information on rejection, approval and negotiation.
Listening to procedural speech	Identifying information about location directions, procedures, procedural agenda, and story plots in the film.

The special characteristic of the listening test indicators and instruments generalized from the opinions of the informants is the purpose of the listening test instruments which requires focusing to compare experiences and knowledge about the recording together. This needs to be compared with the objective of listening that is informative. In the objective of informative listening, test takers are not required to provide a critical interpretation of the recording (stimulus) received. The reaction given by the test takers must have been determined by the researcher from the start.

Time constraint is a factor that the test developer must pay attention to. Test takers were given time in accordance with the number of items determined. In addition, each item to be worked on by a foreign speaker has a pause to understand the next number of question. The pause specified for the listening instruments in this study must be adjusted to the length of the dialog and monologue played. The average pause for working on each number is 7 seconds while the time allocated to review the next number of question is 8 seconds. The total pause given between numbers is 15 seconds.

The factor of various forms of recordings was also considered based on the differences in the characteristics of foreign speakers. Each country certainly has different linguistic rules considering the process of the language acquisition of the foreign speakers. This factor was eventually omitted because it is not related to the mastery of Indonesian language. Foreign speakers must adapt to the use of Indonesian language in everyday communication purposes. The solution to this problem is by developing test items that do not contain differentials, especially regarding language talents, interests, and attitudes of foreign speakers.

The listening test ultimately needs to be classified into four levels, namely the levels of memorizing, comprehension, application, and analysis. Memorizing requires foreign speakers to remember the facts and collect the information being heard(Saricoban, 2012). Understanding requires foreign speakers to understand the content of discourse, inter-event relationships, between relationships. Application is the resolution of problems based on information listened to by foreign speakers. Analysis leads to the consideration of certain linguistic forms and aspects by determining whether the response is appropriate.

Listening skills materials in the Indonesian language proficiency test has two focuses. First, it is part of general speaking skills including dialog and role playing. One of the most difficult, often weak, aspects is the listening material in the context of speaking, especially when speaking in a foreign language that cannot be predicted by the interlocutor. According to Cunningsworth, the most effective strategy is to think quickly and accurately in predicting the interlocutor's speech when listening(Cunningsworth, 1995). Indonesian language proficiency test can help the students to overcome problems in this context. For example, by making an outline of the contents of the dialog, what the students say can be well controlled and assessed. However, it is necessary to design the answers that might be responded to.

Inputs on the needs analysis process is understanding the linguistic feasibility of the test instrument which must be emphasized. This refers to variations in writing styles that are precise and easy to understand, diverse language structures that are not monotonous, according to the ability of foreign speakers (not native speakers), appropriate terms/terminology, and the presentation of information in context. Based on these inputs, it can be concluded that the language of the instruments to be developed must be (1) according to the ability of foreign

*Corresponding Author: Memet Sudaryanto，Email: memet.sudaryanto2016@student.uny.ac.id

speakers, (2) communicative, (3) appropriate, and (4) coherent. The four indicators must be used as a reference in developing item specifications, indicators, test instruments, and descriptions of the competence of foreign speakers.

The dimension factor is the eigenvalue that forms each dimension in the instrument being analyzed. Unidimensional criteria can be seen from the scree plot formed on the distribution of the analysis of the test takers' ability to answer each instrument. Michael Furrexplains that a test is unidimensional if the components 1 and 2 on the scree plot have a long steep distance(Furr, 2014). This is reinforced by the opinion of Retnawatithat the assumption of unidimensionality as well as local independence can be done with factor analysis by examining the outputs on the resulting scree plot and the eigenvalue of the variance explained(Retnawati, 2017). Based on the results of the factor analysis, it was obtained that the response of foreign speakers contained several components that had eigenvalues of more than 1.0, but component 1 had the highest steepness with an eigenvalue of 3.499. Based on the scree plot, it is clear which component has the highest steepness (based on eigenvalues) is the first component.

Conclusion

Foreign speakers in Indonesia take the test that is usually tested on native speakers of Indonesian. One basis for assessment is the conformity of test characteristics for test takers so that item information can explain the test takers' abilities more accurately. Based on this concept, this study develops two measurement products, namely (1) standardized tests for receptive skills of Indonesian language for foreign speakers and (2) competency assessment of foreign speakers who are clustered at three levels of proficiency, which are average, intermediate and advanced.

The reliability estimate shows good results, which is the reliability with IIF shows that the Y line shows the maximum size of the information function. The reliability Cronbach's alpha shows an estimate of 0.700 and the reliability with the construct shows a value of 0.87. Based on the estimated reliability of the instrument, an advanced test can be performed, namely the assumption test using Item Responses Theory. The foreign speakers' assessment standards should be re-determined. The Ministry of Education and Culture through the Language Body has arranged competency assessments into seven standards whose implementation is interpreted differently in each organizer. Research development specifically divides the skills/competency assessment of foreign speakers into three simpler levels and their development through the 1 parameter approach in IRT. The competence of foreign speakers is divided into three levels, namely average, intermediate, advanced for each dimension of the test.

The assessment of foreign speakers' abilities is shown by the results of the implementation of the bookmark method which shows at which level the test taker can master the language skills. The results of the competency assessment show that 39% of the foreign speakers are at the intermediate level, 33% at the advanced level, and 28% at the average level.

References

[1] Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. Educational and Psychological Measurement. https://doi.org/10.1177/0013164485451012

[2] Brown, R. D. (2012). Book Review: Urban Microclimate: Designing the Spaces Between Buildings. Urban Studies. https://doi.org/10.1177/0042098011435914

[3] Cizek, G. J., & Bunch, M. B. (2007). Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Tests. In Sage Publications. https://doi.org/10.1007/s13398-014-0173-7.2

[4] Cunningsworth, A. (1995). Choosing your coursebook. Oxford: Heinemann.

[5] Elder, C., & McNamara, T. (2015). Exploring the limits of authenticity in LSP testing: The case of a specific-purpose language test for health professionals. Language Testing. https://doi.org/10.1177/0265532215607397

[6] Furr, R. M. (2014). Scale Construction and Psychometrics for Social and Personality Psychology. In Scale Construction and Psychometrics for Social and Personality Psychology. https://doi.org/10.4135/9781446287866

[7] Hambleton, R. K., & Swaminathan, H. (1985). Item response theory : principles and applications. In Evaluation in education and human services; https://doi.org/10.1017/CBO9781107415324.004

[8] Khairani, A. Z. bin, & Razak, N. bin A. (2013). Advance in Educational Measurement: A Rasch Model Analysis of Mathematics Proficiency Test. International Journal of Social Science and Humanity. https://doi.org/10.7763/ijssh.2012.v2.104

[9] Kim, D., & Gilman, D. A. (2008). Effects of text, audio, and graphic aids in multimedia instruction for vocabulary learning. Educational Technology and Society.

[10] Retnawati, H. (2008). Validitas dan reliabilitas konstruk skor tes kemampuan calon mahasiswa. (1), 126–135.

360

[11] Retnawati, H. (2017). Proving content validity of self-regulated learning scale (The comparison of Aiken index and expanded Gregory index). Research and Evaluation in Education. https://doi.org/10.21831/reid.v2i2.11029

[12] Saddhono, K. (2015). Integrating culture in Indonesian language learning for foreign speakers at Indonesian universities. Journal of Language and Literature. https://doi.org/10.7813/jll.2015/6-2/58

[13] Saddhono, K., & Rohmadi, M. (2014). A sociolinguistics study on the use of the Javanese language in the learning process in primary schools in Surakarta, Central Java, Indonesia. International Education Studies. https://doi.org/10.5539/ies.v7n6p25

[14] Saricoban, A. (2012). Translation to Test Vocabulary and Language Skills. Procedia - Social and Behavioral Sciences. https://doi.org/10.1016/j.sbspro.2012.05.597

[15] Terrell, T. D., & Brown, H. D. (2006). Principles of Language Learning and Teaching. Language. https://doi.org/10.2307/414380

[16] Winarni, R., Slamet, S. Y., & Saddhono, K. (2018). Development of Indonesian literature textbook with character education through information and communication technology (ICT) learning based. International Journal of Engineering and Technology(UAE). https://doi.org/10.5465/AMR.2009.35713291

*Corresponding Author: Memet Sudaryanto，Email: memet.sudaryanto2016@student.uny.ac.id