

Indonesian as a Foreign Language: Standard Setting and Materials Development Issues

Memet Sudaryanto^{1,2}, Kundharu Saddhono¹, Hari Wahyono³, Tri Widiatmi⁴, La Ino⁵, Herry Nur Hidayat⁶, Asep Yudha Wirajaya¹, Imam Suhardi⁷, Budi Waluyo¹, Pramono⁶

Sebelas Maret University, Surakarta Indonesia¹

Yogyakarta State University, Yogyakarta, Indonesia²

Tidar University, Magelang, Indonesia³

University of Veteran Bangun Nusantara, Sukoharjo Indonesia⁴

Halu Oleo University⁵

Andalas University, Padang Indonesia⁶

Jenderal Soedirman University⁷

{memetsudaryanto@staff.uns.ac.id¹}

Abstract. The research analyzes the standard setting and test material on Indonesian language test, specially for foreign students in Indonesia. The research used qualitative approach by doing in depth interview and content analysis. Data analysis techniques which were used are interactive patterns consisting of data collection, data reduction, conclusions, and data displays. The standard setting analysis was done by in depth interview with Indonesian language instructor for foreign speaker, document analysis (of) in the form of graduate competency standard, test instrument on listening skill, speaking, reading, writing, and grammar, assessment rubric. The needs for test materials design was identified by a questionnaire of material presentation and structural arguments that support judgment from expert. The constructs of test materials are compiled from BIPA textbooks that are compiled into a list of test materials to be assessed by the expert. The expert scored the importance of the material or the basis of the criteria. The results of the study indicate that an accurate standard setting is required in placing test participants at a level appropriate to the competence (ability) and based on the test material to suit the needs of BIPA students.

Keywords: qualitative research, BIPA, Foreign Language.

1. Introduction

One of the steps to ensure the quality of the assessment is the determination of the standard setting as a reference for the success of learning. Standard setting places students to know the improvement of their competence (Alsmadi and Jordan, 2007; Lin, 2006). Student's competence measured by tests and non-test specially designed for the purpose of the assembled instrument. One of the essential functions of setting the standard setting is threshold of the pass need to be analyzed of its position, whether it is representative of

standardized and valid process or not. In Indonesia, the reference standard used so far follows the qualitative reference without considering the analysis of the difficulty level of the questions and competencies of test's participants (Widanarto, Mardapi, Budiyo, 2015: 179). Added by Cizek and Toledo (1996: 14) which explains that the policymaker is responsible for setting the passing grade often refers to curriculum rules, the public interest, the constraints of choice to provide reasonable grounds without accurately calculating the passing limit. Determination of standard setting requires a reference instrument in accordance with the ability of test participants with test materials that have been designed according to the needs of test participants.

There are at least two ways of setting standards that can be used as reference for teachers and policy makers. Retnawati (2014: 165) describes two methods of setting the standard setting in a test, the standard setting based on the grain / test and standard setting designed based on the test participants. The setting of the standard setting has different procedures. Unlike Anto and Mardapi (2013: 370) which classify the standard setting in two types, (1) standard setting using norm reference, and (2) standard setting with reference criteria. Based on the results of interviews with BIPA teachers, currently the determination of standard setting for BIPA student competence has not been quantified according to the correct procedure. One of the standards used to determine graduation criteria also adopts CEFR's (The Common European Framework of Reference for Languages). Various European institutions have also set up their own standards, which are more or less related to those produced by the Committee (Dasheng and Varghese, 2013).

On the other hand, the unpreparedness of determining standard setting is supported by an inadequate Indonesian language skills test instrument that accurately measures the competence of foreign students. The UKBI instrument assessment by Ariamaharani (2017: 47) consists of 5 different measured competencies, namely listening skills, speaking skills, reading skills, writing skills, and the ability to respond to the rules. The Indonesian test is specified into listening, speaking, reading, writing, vocabulary, grammar, pronunciation, and up to the punctuation use (Dasheng and Varghese, 2013). Brooks (1964: 134) classifies language skills into receptive and productive skills. Receptive skills consist of listening, reading, and competence respond to rules. The competence of speaking and writing belongs to productive skills. Receptive skills are measured by multiple choice tests, while productive skills are measured by performance.

The index of difficulty items on UKBI does not match the competency level of the test participants, in this case the BIPA students. Sumintono and Widhiarso (2015: 25) explained that a good item is what matches the ability of the test participants to be measured. The characteristics of the items based on the IRT concept consist of one, two, and three logistic parameters. The parameter component consists of (b) parameter of the index of difficulty items, (a) index of discriminant parameters, and (c) pseudo guessing parameters. At least the item has an index of difficulty item parameter to determine the ability of test participants to answer correctly. The grain-level difficulty parameter is a point on the scale of the ability to correctly answer the probability of 50%, because the greater the value of the index of difficulty it is necessary to answer correctly as much as the 50% chance (Retnawati, 2014; Cizek and Toledo, 1996).

The standard setting currently used by the language department refers to the UKBI instrument (Indonesian Language proficiency test) in which the test is intended for all test takers who want to know the standard of their Indonesian language skills. BIPA assessment standards still require a lot of evaluation from the preparation of materials and the ability of test participants (Ariamaharani, 2017: 48). Internal testing conducted in the teaching of BIPA

according to Maryanto (2001: 390) can be considered not enough to evaluate the capabilities of foreign speakers of bahasa Indonesia. Beside evaluation materials are now used in BIPA's teaching is still very varied, BIPA's teaching is only one activity that allows foreign speakers learning the language. The determination of the difficulty index in IRT is analyzed based on the ability of the test participants, so that the characteristics of different test sample population will produce different item characteristics and tend not to be identical. On the other hand, setting the standard setting requires the information of the difficulty index parameter to determine the passing threshold, as well as the standard level achieved by the test participants.

2. Literature Review

The flow of problems in this article are (1) the validity of UKBI instrument test materials based on the subject, i.e. BIPA students, (2) difficulty index parameters on UKBI instruments based on the population of the research subjects, (3) the teacher's perspective on setting standards and (4) standard setting techniques used by teachers of BIPA are procedural and tested, (5) determination of standard setting partially on each language skill and determination of total score of pass limit, and (6) descriptor each level that determines the graduation of BIPA students. This article will generally address the flow of the problem in two main formulas namely (1) perception of BIPA lecturers' needs regarding standard setting and quantitative procedures of determination, and (2) preparation of test materials in accordance with the needs of BIPA students and their contextual testing. These two formulas will be the fundamental policy-making that needs to be analyzed.

2.1. Indonesian Language Skills

Assessment of language skills requires different assessment techniques. Listening skills is the ability to respond to speech by determining the most appropriate response to the information obtained. Speech skills can not be measured by dichotomous scans, assessment techniques on these skills require both partial and graded performance and performance. Assessment of reading skills as receptive skills requires receptive assessment techniques such as listening skills. Finally, writing skills with data retrieval techniques in the form of productive skills have similarities with the speaking skills of politomus judgments.

Language skills as a primary requirement of communication require a standard of survival skills as part of a speech-sharing culture. Language skills consist of four skills, each of which is integrated as a reciprocal relationship. In the context of this study, (1) the subject of research needs are foreign students studying in Indonesia, (2) the language skills materials tested are listening, speaking, reading, writing, and grammar skills as a form of response rule. Brown (2003: 119) reveals the question, "how can one speak without listening", Brown added that the assessment of listening success is not rated equivalent to speaking skills. Listening is a process that involves listening to the sounds of language, identifying, interpreting, valuing, and reacting to the meaning contained in oral discourse.

Listening is defined by Tarigan (1987: 28) as a process of listening to oral symbols attentively, appreciation comprehension, and interpretation to obtain information, capture content or messages and understand the meaning of communication that has been conveyed by the speaker through speech or spoken language. A strong correlation between listening and speaking, namely (1) a speech obtained by a person / child through listening followed by imitating the way of pronunciation. The utterance or vocabulary mastery will be used or used

in speaking; (2) both of these skills require good cooperation. When two people communicate, they can not do it simultaneously, both speaking or both listening. This activity should be done alternately (reciprocal), when one speaks, the other listens or listens attentively and vice versa. Sutari et al. (1998: 4) states more aspects of the language skills involved in listening activities, the better the outcome.

Michael Rost (1991: 4-5) writes that a listener should have the following abilities: (a) Distinguish sounds, (b) form syllables into words, (c) identify word groups, (d) identify pragmatic elements, such as expression, companion, place, time, and purpose, (e) consider linguistic and paralinguistic aspects (intonation or pressure) and other aspects beyond linguistic, (f) utilize the knowledge already possessed which relates to the content of the utterance being observed so as to predict and capture meaning appropriately, (g) to understand the expressed or implied words or ideas or main ideas. Rost further explains that someone is said to be successful in listening if he is able to connect / use these abilities. Abilities above are grouped into three as the ability to understand, analyze, and identify.

Assessment of the ability to speak in the teaching of language based on two factors namely linguistic and non-linguistic factors. Linguistic factors include pronunciation, vocabulary, and structure while non-adaptation factors include material, fluency and style (Haryadi, 1997: 95). Speaking as a productive skill has product characteristics that can be observed by the assessor. Speaking is one of the most difficult skills that language learners have to face. In spite of this, it has traditionally been forced into the background while we, teachers of English, have spent all our classroom time trying to teach our students how to write, to read and sometimes even to listen in a second level study because grammar has a long written tradition (Bueno, Madrid and McLaren, 2006: 321). For most people, the ability to speak a language is synonymous with knowing that language. Nevertheless, speaking in a second or foreign language has often been viewed as the most demanding of the four skills.

Reading is a simple process because it does not require as much equipment as listening. Reading is able to improve the insight of knowledge and knowledge of learners. The important thing in reading is to capture the idea, not to remember or ponder the symbols printed therein (Soedarso 2002: 5). Added by Hernowo (2005) the most common reading benefit is to obtain information and knowledge, while the specific benefit of reading activities is increasing the power of brain function. Reading text can be used for a number of purposes, among others, to develop students' reading skills, show good grammar and correctness, increase vocabulary mastery, illustrate different writing styles, provide interesting information, and stimulate speech or other skills (Cunningsworth, 1995: 73). His vocabulary, interest, eye reach, speed of interpretation, previous experience background, intellectual ability, familiarity with read ideas, reading goals, and flexibility set the pace are the determinants of reading success (Soedarso 2002: 58).

2.2. Standard Setting Selection

The policymaker responsible for establishing the passing grade often refers to the legislative, general, optional or prior referral mandates without modification and provides reasonable grounds for determining the passing limit of the test taker regardless of the testers' ability. Mardapi, et al (2015: 40) mentions three popular techniques in determining the minimum mastery, namely judgment, tests, and test participants. Standard setting is developed to make reference to the competencies that the test takers are capable of. Setting the standard setting is a form of quality improvement or educational standard that is viewed from the perspective of test participants. Khatimin (2013: 151) explains that setting standard setting is

used to ensure that the performance of the test participants meets the criteria specified as the limit of the pass.

The use of the Council of Europe Framework of Reference (CEFR) is considered less effective. Tests conducted in Indonesia to measure the competence of Indonesian language with European assessment standards. Standards that were not adapted without prior research (Kenyon, 2013: 1) were considered unacceptable decisions. The Garcia, et al (2013: 243) study suggests that linking language skills to CEFR involves setting cut scores that use classification based on pre-defined frames. Lissitz and Li (2011: 476) divides the standard setting level into 4 groups, namely (1) advanced, i.e. students who are capable of exceeding the expected ability to demonstrate independence and an accurate understanding of the concepts and skills tested. Students can demonstrate their ability to apply skills / concepts to authentic tasks, (2) proficient, i.e. students who can demonstrate independence and understanding accurately although generally there is no conformity with conceptual understanding that needs to be demonstrated, (3) basic, i.e. students show basic understanding of specific skills, and (4) students showing little or no understanding of the skills / concepts being taught. The explanation supported by Khairani, Razak, and Shamsuddin (2014: 198) who use three predetermined cut scores to differentiate students with different levels of performance, which are basic, proficient, and advanced. The results show that panelists have high satisfaction in terms of practicality procedures and application procedures.

3. Research Method

This research is a qualitative research. The methods of data collection are observation and interview techniques. The data sources of this research are (1) Indonesian textbooks for foreign speakers (BIPA), (2) BIPA teachers who have been teaching for more than 5 years and have developed their own teaching materials, (3) BIPA students who have graduated and use Bahasa Indonesia as one of the skills, (4) the BIPA exam instrument used by the teacher, and (5) the student scoring document after passing the BIPA learning.

The research was conducted at BIPA organizers in Yogyakarta, Surakarta, Semarang, and Surabaya. The lecturer who became the resource person in this research has several criteria, (1) the BIPA teachers are the speakers of BIPA, (2) have teaching experience and tutor BIPA within 5 years, (3) develop BIPA teaching materials for foreign students in Indonesia/abroad studying Indonesian language, (4) developing BIPA learning assessment instrument (four language skills). Development of interview guides based on mapped needs analysis to determine the competency standards of graduates and perceptions of BIPA teachers on the use of standard setting as a reference for graduating BIPA students.

4. Result and Discussion

4.1. Language Proficiency Setting Standard

Types of tests used by several institutes of the program of Bahasa Indonesia for Foreign Speakers (BIPA) are different, i.e. using placement test or test proficiency. Placement test is used to see the most appropriate level based on qualitative observation. Unfortunately, from testing four language skills, the newly used test tool measures writing and speaking skills with written tests and interviews. In contrast to the proficiency test used in some organizers only

measure listening, reading, and grammar skills. Testing of proficiency test using instrument of UKBI where the difficulty level of item is not balance with ability of test participant.

The next stage, the test is used to place test takers at a level appropriate to their competence. Progress tests use a range of values consisting of A (Perfect), B (Very Good), C (Good), D (Enough), and E (Less). For the test, each teacher has different assessment techniques. Talking about standard setting issues, some BIPA providers use test tools to place test participants at the specified level. Determination of BIPA test participants level uses CEFR reference with 6 standards. Though it should use language body references, some institutions find it difficult (1) the division of classes into seven levels requires substantial fees for teacher fees and class divisions, (2) material of language level has not mapped well, and (3) descriptor at each level sometimes has the same test standard. When placement test, because the body of the language specification is unclear and concrete, then use CEFR with clear descriptor and adapted to be its own reference. When placement tests, CEFR places students with written tests and speak as productive skills. Listening skills are tested on a progress test, there is a value range category (listening to audio). For example (1) Interview: appropriate response, inappropriate, fatal error, (2) listening vocabulary, based on simakan, (3) test listening to retell. The basic level for the simplest listening is the vocabulary test with the cloze test. This standard is created with a range of values. Standards by teacher judgment but converted by management standards for certification of examinees.

4.2. Determination of BIPA student passing standards

The determination of the BIPA students passing limits on various campuses using different methods, namely (1) multiple cut score with criterion B or good, just passed on this criteria students with limit C, D, and E are not declared pass before follow remedy, and (2) single cutscore which is 70. The teacher is not aware of the standard setting. Teachers are briefed on the use of setting standards previously determined by the policy makers, in other words live applications on the level already provided and the curriculum provided by the university / BIPA organizers. Retnawati (2014: 171) explains that the bookmark procedure is based on IRTs that combine simultaneously between the characteristics of the testers' abilities and the difficulty of the items. On the other hand, Mardapi et al. (2015: 40) describes eight panelist determination requirements to become a field expert, (1) best in his specialty field, (2) have extensive insights in his area of expertise, (3) have ability to solve the problem fast according to the expertise (4) able to deeply examine the conceptual level of the field rather than the new person, (5) analyze the problems in the field qualitatively, (6) assess the problem more accurately than the new person, and (7) have more complex semantic memory.

Descriptors of each level (A, B, C, D, E) have teacher report results and managers determine at the level. In the test participants with the highest level primarily students with academic goals, meaning lecture planning, credit transfer program. At this level, test participants can already interact very well. Most students are expected to enter the minimum level B.

Level A	Can master the Indonesian language and apply in the context of everyday communication.
Level B	Vocabulary mastery has been in accordance with the field of science that was cultivated although limited to the number of words that are still common and easily understood
Level C	Test participants can use everyday expressions and vocabulary with specific contexts.

Level D	Patterns are mainly on receptive skills and the test participants are still misunderstood but not fatal.
Level E	The test participants in composing the sentence structure can not be understood

5. Conclusion

In general, it is necessary to rearrange the competency standards of BIPA graduates in accordance with the standards that are more acceptable and in accordance with the level of testers' ability. Ability of the test participants starting from zero skills (test takers not mastering Indonesian language at all) to the highest skills (the skills of test participants to use Indonesian as a colloquial language) Based on the need assessment conducted with BIPA teachers, a test is needed to place test participants at their competency level.

Some BIPA operators use test kits to place the test takers at the specified level. Determining the level of BIPA test participants using CEFR reference, with 3 standards to facilitate BIPA test participants and teachers in determining the learning treatment that will be given. Each language skill (listening, speaking, reading, writing, and grammar) has a set of standard/standard setting limits, for a standard setting total of four language skills averaged proportionally based on the principle of need. It is necessary to determine the boundaries of BIPA students passing on various campuses using the same method, that is, by multiple cut score according to new criteria to determine the test participants passed on this criterion of students with a certain limit.

References

- [1] Alsmadi, A.A. (2007). A Comparative study of two standard setting techniques. *Social Behavior and Personality Journal*. 35 (4). Hlm. 479-486
- [2] Anto, S. and Mardapi, D. (2013). Komparasi Metode Standard Setting Untuk Penentuan Kkm Mata Pelajaran Matematika Kelas VIII SMP. *Jurnal Penelitian dan Evaluasi Pendidikan*. Vol.17 No.2 hlm. 369-388
- [3] Ariamaharani, Andhika. (2017). Pengembangan Alat Evaluasi Pembelajaran BIPA. *Simposium Internasional Pengajaran Bahasa Indonesia bagi Penutur Asing (BIPA)*. Pusat Pengembangan Strategi dan Diplomasi Kebahasaan. Yogyakarta, 23—24 Agustus 2017. Hlm.46-51
- [4] Brooks, Nelson. (1964). *Language and Language Learning*. New York: Harcourt, Brace and World, Inc
- [5] Brown, H. Douglas. (2003). *Language Assessment Principles and Classroom Practices*. USA: Longman
- [6] Bueno, A, D. Madrid and N. McLaren, (eds). (2006) *TEFL in Secondary Education*. Granada: Editorial Universidad de Granada.
- [7] Cizek, G.J. (1996). *Standard Setting Guidelines*. *Educational Measurement: Issues and Practise*. Hlm.13-23

- [8] Cunningsworth, Alan. (1995). *Choosing your Course Book*. Oxford: Heinemann
- [9] Haryadi. (1997). *Berbicara (Suatu Pengantar) Diklat Perkuliahan*: IKIP Yogyakarta.
- [10] Heri Retnawati. 2014. *Teori Respon Butir dan Penerapannya*. Yogyakarta: Parama Publishing.
- [11] Heri Retnawati. 2016. *Validitas dan Reliabilitas Karakteristik Butir Soal*. Yogyakarta: Parama Publishing.
- [12] Hernowo. (2005). *Quantum Reading*. Bandung : MLC
- [13] Maryanto. (2001). 'Tes UKBI dan Pengajaran BIPA'. 4th International Conference on the Teaching of Indonesian to Speakers of Other Languages. KIPBIPA IV Indonesia Australia Language Foundation. Hlm.379-391
- [14] Rost, Michael. (1976). *Listening in Action Activities for Developing Listening in Teaching*. New York: Prentice Hall International (UK) Ltd.
- [15] Sumintono, B dan Widhiarso, W. (2013). *Aplikasi Model Rasch untuk Penelitian Ilmu-Ilmu Sosial*. Cimahi: Trim Komunikata Publishing House
- [16] Sutari, Ice, dkk. (1997). *Dasar-dasar Kemampuan Menulis*. Bandung: FPBS IKIP Bandung.
- [17] Tarigan, H.G. (1986). *Menyimak sebagai Suatu Keterampilan Berbahasa*. Bandung Angkasa
- [18] Widanarto, S.P., Mardapi, D., Budiyo. (2015). Perbandingan Estimasi Kesalahan Pengukuran Standard Setting dalam Penilaian Kompetensi Akuntansi SMK. *Jurnal Penelitian dan Evaluasi Pendidikan* Vol.19 No.2 hlm.176-188
- [19] Khatimin, N., Aziz, A.A., Zaharim, A. and Yasin S.H.M. (2013). Development of Objective Standard Setting Using Rasch Measurement Model in Malaysian Institution of Higher Learning. *International Education Studies*. Vol. 6 (6) hlm. 151-161
- [20] Mardapi, D., Hadi, S., dan Retnawati, H. (2015). Menentukan Kriteria Ketuntasan Minimal Berbasis Peserta Didik. *Jurnal Penelitian dan Evaluasi Pendidikan* Vol.19 No.1 hlm.38-45
- [21] Garcia, P.E., Abad, F.J. and Aguado, D. (2013). A New IRT-Based Standard Setting Method: Application to eCat Listening. *Psicothema* Vol.25 no.2 hlm.238-244
- [22] Kenyon, Dorry. (2013). *The Encyclopedia of Applied Linguistic*. Edited by A. Chapelle. Blackwell: Blackwell Publishing Ltd.

[23] Lissitz, R.W., and Li, F. (2011). Standards Setting in complex performance assessment: An approach aligned with cognitive diagnostic models. *Psychological Test and Assessment Modeling*. Volume 53 No.4 Hlm.461-485

[24] Khairani, A.Z., Razak, N.A., Shamsuddin, H. (2014). 'Application of the Rasch Model and the Bookmark Method in Setting Cut Scores in Mathematics'. *International Journal of Information and Education Technology* Vol.4 No.2 April 2014