

THE MODEL OF BIPA LISTENING EVALUATION AND ITS IMPLICATIONS FOR THE DESIGN OF LISTENING EVALUATION

Ida Widia, Syihabuddin, Vismaia S. Damaianti, and Yeti Mulyati
Universitas Pendidikan Indonesia

*Corresponding author: idawidia@upi.edu

First draft received: 1 Sep 2021

Date Accepted: 15 Feb 2022

Final proof received: 28 Feb 2022

Abstract

In this paper, an analysis of the results of the evaluation of Indonesian language proficiency for foreign speakers is presented. The development of a standardized Indonesian language for foreign speakers (*Bahasa Indonesia bagi Penutur Asing* - BIPA) evaluation tool is currently considered very important. Currently, the existing evaluation tools are still made partially depending on the needs of institutions that provide the Indonesian language program for foreign speakers (BIPA). In addition, this need is also driven by the need for foreign speakers to measure their proficiency in Indonesian. Although there is an Indonesian Language Proficiency Test (UKBI) as an evaluation tool in Indonesia, UKBI is considered not in accordance with the need to measure the Indonesian language proficiency of foreign speakers because UKBI is still used to measure the Indonesian language skills of native speakers. This has become one of the problems in the field of evaluation of the Indonesian language, especially BIPA. The language evaluation tool used to measure a person's foreign language proficiency always begins with a listening proficiency test. Listening is considered the gateway to other language proficiency. The solution to this problem is to analyze listening evaluation tools in foreign languages that are used continuously, such as the Test of English for International Communication (TOIEC), the Japanese-Language Proficiency Test (JLPT), and *Diplôme d'études en langue française* (DELFL), which represents three continents; American, Asian, and European spoken by Indonesian speakers. This study used a descriptive method to solve this research problem. The results of this study describe the profile of the listening proficiency evaluation tool in foreign and Indonesian languages in three ways: 1) speakers (speaker stimulus presenter), 2) content, and 3) audio. Furthermore, the results of this analysis can also be used as the basis for developing the BIPA listening competency evaluation model.

Keywords: Evaluation language; foreign speakers; listening competence.

To cite this paper (in APA style):

Widia, I., Syihabuddin, Damaianti, V. S., & Mulyati, Y. (2021). The model of BIPA listening evaluation and its implications for the design of listening evaluation. *Indonesian Journal of Education*, 15(1), 28-37. <https://doi.org/10.17509/ije.v15i1.46154>

INTRODUCTION

Indonesian language for foreign speakers or BIPA (*Bahasa Indonesia bagi Penutur Asing*) often gets great attention from various parties, not only from academics and BIPA teaching practitioners, but also from government, private parties, and other parties with interest in developing Indonesian language ability of foreign speakers. Currently, the BIPA program has developed rapidly both in Indonesia and abroad. Approximately 219 BIPA organizing institutions are spread across 74 countries (Wahya cited in Muliastuti, 2017).

Mastery of the Indonesian language as a communicative foreign language is considered capable of supporting the progress of other sectors such as business, tourism, diplomacy, and others. These interests are directly related to the social life of Indonesian society. Language proficiency and mastering Indonesian are important to learn so that foreign speakers can communicate in communicative Indonesian (Yeyen, 2017).

Like teaching foreign languages, in general, the rapid development of BIPA learning requires the readiness of standardized BIPA learning tools in order to be able to run optimally. As Alwi (2000) stated, apart from standardizing the learning tools for the BIPA learning program, its evaluation tools also need to be standardized.

The evaluation of the Indonesian language as a foreign language is still being partially developed by each institution that organizes Indonesian language programs or courses in accordance with the learner's needs. However, until now, no evaluation tool has been found capable of measuring standardized BIPA language proficiency. This was confirmed by Mulyati (2006), stating that every foreign language institution should have standardized test kits to evaluate after learning.

As an institution that protects the development and integrity of the Indonesian language, the Language Agency states that one of the efforts to maintain language is to create a means of measuring Indonesian language proficiency, namely the

Indonesian Language Proficiency Test (UKBI/*Uji Kemahiran Berbahasa Indonesia*). It has a strategic function, not only to improve the quality of the Indonesian language as well as its use and teaching but also to foster a positive attitude and a sense of pride in the Indonesian people towards their language. However, fostering this positive attitude also applies to foreign speakers who want to measure language proficiency so that some things are less relevant to BIPA teaching that puts forward a communication approach.

In fact, UKBI is still used to test the language proficiency of Indonesian and non-Indonesian speakers. Several interviews with foreign speakers who participated in UKBI stated that they had difficulty following the listening evaluation pattern whose problem placement was not in order. However, UKBI only provides a standard assessment of Indonesian language users' ability without considering the differences in their learning situation, whether foreign speakers or native speakers are treated the same (Tsamaratul, 2011). Tsamaratul (2011) also revealed that UKBI had not been tested to measure the level of Indonesian language proficiency for foreign speakers because UKBI questions are more aimed at Indonesian native speakers. Hence, the measuring instrument used should be different. If UKBI is used to measure Indonesian language skills for Indonesian native speakers, then there must also be a measuring instrument that is in accordance with the level of BIPA learners' ability (Ministry of Education and Culture, 2017). Fulcher (2014) explains that testing and assessing foreign language speakers is an important part of the learning process.

In the process of evaluating foreign languages, it was found that listening proficiency was a skill that was always being tested. This is in line with Hubackova's opinion (Effendy, 2006) that listening evaluation is one language skill that receives important attention from all foreign language institutions. The listening evaluation of foreign languages must be considered because foreign speakers still have difficulty listening to Indonesian.

This is evidenced by Castro et al. (2015) that failing to comprehend Indonesian language listening skills is more critical than failing to learn writing and reading skills because the learners are not native speakers. In listening, foreign speakers must distinguish sounds, interpret tonal stress, understand vocabulary and grammatical structures, infer them in a sociocultural context, and relate information to a communicative context (Sugiyono, 2009).

Based on the problems mentioned above, the formulation of the problem in this study is "How is the profile of listening competence evaluation tool of the existing foreign language?" In general, this study aims to show the result of the study of the listening competence evaluation tool of the existing foreign languages in use today and discover the model design of the listening competence evaluation tool used in evaluating the competence of foreign speakers.

THEORETICAL FRAMEWORK

Language evaluation is a measurement activity causing a person to try to improve his initial ability based on the results of the evaluation that is followed. The evaluation means assessment, interpretation, consideration, or judgment (Sugiyono, 2009) of the result of processing and determining the measurement of things or objects based on certain references in determining certain objectives. This definition of evaluation applies to all fields to measure the competence objectives' achievement, including in teaching the Indonesian language as a foreign language. Until now, there has not been found an evaluation tool for the Indonesian language as a foreign language with international standards such as the Test of English for International Communication (TOEIC), Japanese-Language Proficiency Test (JLPT), and *Diplôme d'études en langue française* or Diploma in French Language Studies (DELFL) used to measure the foreign language proficiency of foreign speakers or target speakers wherever they are simultaneously and with the same form and question items. In Indonesia, currently, measuring language proficiency and Indonesian language for foreign speakers is still using UKBI. In comparison, UKBI itself is still used to measure the Indonesian language proficiency of Indonesian native speakers.

Listening competence evaluation receives important attention due to the fact that currently, foreign speakers still have difficulty listening to the Indonesian language. This is evidenced by Castro (2015) that the failure to comprehend Indonesian language listening skills is more critical than the failure to understand writing and reading skills because the learners are not native speakers.

Listening Competency Test Construction

As with all effective tests, designing appropriate assessment tasks in listening begins with the specification of objectives or criteria. Those objectives may be classified in terms of several types of listening performance "Think about what you do when you listen." Literally, in nanoseconds, the following processes flash through your brain.

1. You recognize speech sounds and hold a temporary "imprint" in short-term memory.
2. You simultaneously determine the type of speech event (monologue, interpersonal dialogue, transactional dialogue) that is being processed and attend to its context (who the speaker is, location, purpose) and the content of the message.
3. You use (bottom-up) linguistic decoding skills and/or (top-down) background schemata to bring a plausible interpretation to the message and assign a literal and intended meaning to the utterance.
4. In most cases (except for repetition tasks, which involve short-term memory only), you delete the exact linguistic form in which the message was originally received in favor of

conceptually retaining important or relevant information in long-term memory.

Each of these stages represents a potential assessment objective:

1. comprehending surface structure elements such as phonemes, words, intonation, or a grammatical category
2. understanding of the pragmatic context
3. determining the meaning of auditory input
4. developing the gist, a global or comprehensive understanding.

From these stages, we can derive four commonly identified listening performance types, each comprising a category within which to consider assessment tasks and procedures.

1. Intensive. Listening for the perception of the components (phonemes, words, intonation, discourse markers, etc.) of a larger stretch of language.
2. Responsive. Listening to a relatively short stretch of language (a greeting, question, command, comprehension check, etc.) in order to make an equally short response (Brumfit, 1987).

METHOD

This study used the descriptive method. This method is suitable for answering research questions focusing on analyzing the profile of an international standardized listening competence evaluation tool for foreign languages. Furthermore, the result of the analysis was conducted from the concepts and data obtained to be used as a fundamental study for developing a hypothetical model of listening competence evaluation tool model design designed as a communicative approach for speakers.

The data were collected using the documentation study collection technique to review some of the needed documents in this study, interview techniques for users of listening competence evaluation tools in foreign and Indonesian languages, and questionnaires to collect data about the impact of using standardized listening competence evaluation tools for foreign speakers.

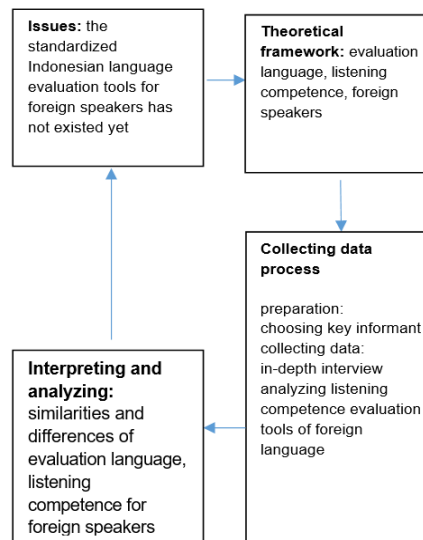
The data sources were the standardized foreign languages listening evaluation tools such as TOEIC, JLPT, and DELF, and Indonesian language listening evaluation tools such as UKBI and TEB (Language Evaluation Test/*Tes Evaluasi Bahasa*), Permendikbud Number 27 the Year 2020, CEFR (Common Europe Framework Research), foreign speakers who took the UKBI or TEB test, Indonesian speakers who took foreign language tests, and Indonesian language teachers for foreign speakers

Research Procedure

An in-depth interview was administered as it is able to dig in information about the main focus and topic of the study deeply, openly, and freely (Effendy, 2006). In this study, the interview was carried out based on pre-

set questions. The procedure of the study is shown in Figure 1.

Figure 1
Research Procedure



The result of interviews and analysis of listening competence evaluation tools were used to identify the profile of evaluation theory-based listening competence evaluation tools in foreign languages and the Indonesian language. Some of the referred theories are those related to listening competence and foreign speakers (Tsamaratul, 2011).

RESULTS AND DISCUSSION

The Profile of Indonesian Language Listening Competence Evaluation Tool

The profile of the Indonesian language listening competence evaluation tool is explained in several aspects: 1) Speaker (presenter stimulus speaker), 2) Content and 3) Audio.

Speaker (Presenter Stimulus Speaker)

The indicators analyzed in the aspect of the speaker or presenter stimulus speaker consisted of 1) pronunciation, 2) intonation, 3) vocal, 4) expression, and 5) accuracy of pause/idea unit. Furthermore, the analysis description from the speaker aspect is as follows.

1) Pronunciation

Pronunciation is one of the indicators on the speaker's part. In UKBI, the pronunciation aspect was good, UKBI listening had already paid attention to pronunciation in the text that was read. The appropriateness of word pronunciation did not change the meaning of the word, so the test takers would easily grasp the meaning of the information conveyed. In the TEB listening, the

pronunciation used by the narrator to convey information was clear. However, there was the pronunciation of a sound accompanied by a strong blowing of air so that the sound [h] was heard in the audio listening. Chaer (2015) explains that the aspiration process is a process in which voiceless consonants are pronounced, followed by a loud blowing out. Based on both results, it was found that both UKBI and TEB had clear pronunciation.

2) Intonation

In the UKBI test, the tempo of the delivered speech was quite fast. The condition of each audio had a different intonation. The first audio had a good tempo and was easy to understand, but the longer the audio was played, the faster the tempo. The speaker's intonation was clear, but the plausible given from moving one dialogue to another monologue was experienced too fast. No regional accent was heard in speech acts that occurred in the audio, so it was not too difficult for the listeners or test takers. Meanwhile, in the TEB test, intonation spoke fast. A good speaking rate is 130 to 165 words per minute (Ministry of Education and Culture, 2017). To avoid misunderstanding between the speaker and the speech partner (listener), the speaker, in this case, was the narrator who was able to pronounce every word clearly. The narrator's intonation was good, with clear articulation and moderate speech plausible. The narrator's speaking intonation when asking questions was clear, and the regional dialect was not heard from the way it was delivered. The speaking intonation when the narrator gave instructions on how to work on the problem in each part of the reading tended to get faster. Although participants in the BIPA 1 learning evaluation test could read the instructions for solving the questions through the question sheets, it would be nice if the narrator's speaking intonation was stable in each part.

3) Vocal

The vocal in the listening session of UKBI test was very clear because the pronunciation and loudness of the narrator voice-enabled the listeners to clearly hear the information conveyed. The vocal in the listening session for the Indonesian Language Learning Evaluation Test for Foreign Speakers Level 1 was clearly heard because the narrator paid attention to the pronunciation and a loud and clear voice when the information was presented. Hence, listeners or examinees could hear clearly, and the information was easy to understand.

4) Expression

The UKBI test, which has four types of dialogue and four types of monologues, contained various expressions such as sadness, worry, happiness, and communicative expressions. The tone colors expressed by the narrator's voice during dialogue or monologue were good and could help listeners understand the condition of the related information.

In line with this, the expressions contained in Indonesian Language Learning Evaluation Test for Foreign Speakers Level 1 could not be seen from the narrator's facial expressions but from listening to the intonation of the speech conveyed. In this Level 1, the dialogues that were delivered contained various expressions such as worried, happy, and expressive. The dialogue was spoken using phrases, words, and expressions related to topics close to life (for example, basic information related to personal and family, shopping, local geography, and work).

5) Accuracy of pause/ idea units

At the same time, the delivery of UKBI test was good by paying attention to the aspects of the accuracy of the pause/idea unit and equipped with several principles of reading aloud, such as the use of intonation, pronunciation, punctuation, and a loud and clear voice when listening to information was conveyed (Brumfit, 1987). This principle was a good indicator when reading aloud, more precisely in the part of the communicator conveying information to the communicant (UKBI test participants).

In the session of listening to the Indonesian Learning Evaluation Test for Foreign Speakers Level 1, the delivery of narrator in conveying information and dialogue was good by paid attention to the accuracy of pause and equipped with several principles of reading aloud, such as the use of intonation, pronunciation, punctuation, and a loud and clear voice when the information was presented (Brumfit, 1987). Given that the listener or test taker was a BIPA level 1 student, the narrator had paid attention to the syntactic aspects well.

Content

The indicators analyzed in the content aspect consisted of 1) situation/register-based language variations, 2) language use, and 3) information delivery. Furthermore, the analysis description of the content aspect is as follows:

1) Situation/register-based language variation

The listening conveyed on the UKBI test was four dialogues with the following details. Two of the four dialogues of the listening session

had shown clearly in the location and situation that occurred on the recording. This was due to the additional back sound, which was quite clear when played. The additional back sound was in the form of birds singing and vehicle sound; for the other two dialogues, there was no additional back sound. This made the atmosphere built into the conversation less clearly defined.

Furthermore, for the monologue session, the four monologues were good. The existence of additional back sound made the situation in the monologue more alive. The additional sounds were in the form of ladies' conversations, crying babies, laughing babies, birdsong, talking people, and instrumental sounds or songs commonly in formal situations such as news reading. This means that a situation would determine the form of language used by the user of that language and its selection based on their respective social conventions. Whereas in the TEB listening, the dialogue conducted by the communicant and communicator was not known for the location of the conversation, because this listening did not use any back sound indicating its location as recorded by other language tests.

The two listenings on the UKBI and TEB tests had significant differences. The UKBI listening test had a situation that tended to be real with the presence of backsound, so that it could make the situation more lively, while TEB listening test did not have any backsound on its listening section.

2) Language use

The language variation user is one part of the speech. This has to do with who the speakers are and where the speech is used. Chaer and Agustina (2004) explain that there are several kinds of languages, including standard, formal, business, casual, and familiar variety language.

On the UKBI test, there were four dialogues and four monologues. In the first dialogue, a woman and man were chatting casually waiting for the herbalist to pass. This indicated that the dialogue was included in the casual mode. The second dialogue discussed child development. The communicator and communicant were married couples. The situation depicted on the listening session was familiar variety language. The third dialogue discussed natural masks and its listening was included a consultative or business variety language. The fourth dialogue explained about Virgin Coconut Oil, the listening was included in the casual language variation.

The first part of the monologue explaining polio was included in the relaxed mode because the communicator explained it in the environment to the communicant who was familiar with the communicator. The second to fourth monologues were included in the formal language variation. The communicator conveyed information in formal situations in the form of news reading and the information delivery at the seminar.

Listening to Indonesian Learning Evaluation Test for Foreign Speakers Level 1 is divided into three sessions. When viewed from the situation of the speech, the three listening sessions were conducted by two communicants and communicators who were the same age and worked as students, although several situations were conducted by discussing family members such as brothers and sisters. Therefore, the dialogue conducted in the listening sessions tended to have a casual and familiar language variation, because it was conducted in informal situations.

Based on the UKBI and TEB tests, both have a casual language variation in the listening. There were also those who used the casual style or the familiar language variation in listening to the audio delivered by the narrator, especially on TEB.

3) Delivery of information

In UKBI, the information was well-delivered and then could be received clearly by listeners. Both were in the form of dialogue and monologue. In dialogue, the delivered information and responses were easy to understand because they had applied the two-way principle, namely, the sender issues ideas and the recipient responds to the content. In other words, there was two-way communication in the form of reciprocity from the communicator and the communicant. Effendy (2006) claim that with this principle that the dialogue that was conveyed on the listening session could be easily accepted by listeners or participants taking the UKBI test.

Likewise, listening to monologues applied for the one-way principle of communication that took place from one party only, namely only from the communicator side by not giving the communicant the opportunity to give a response (Sungkono, 2015).

Meanwhile, the information from listening to Language Evaluation Test for Foreign Speakers Level 1 was clear and accurate because it was delivered directly by the source of the message (dialogue), which could also be responded to by the recipient or, in this position, the test takers. That way, the test takers or the message recipients

could immediately confirm the message they got by answering the listening questions correctly because the used conversation text was in accordance with the BIPA level 1 topic and material. The listening and response dialogue conveyed was easy to understand because it applied the two-way principle; the sender issued ideas, and the recipient responded to the content. In other words, two-way communication occurs in the form of reciprocity between the communicator and the communicant (Effendy, 2006).

The information conveyed in the UKBI test, and the Language Evaluation Test had in common; namely, the delivery of information was good so that the recipient or communicant could clearly understand what the narrator conveyed in the audio listening.

Audio (Recording)

The indicators analyzed in the audio aspect consisted of: 1) clarity, 2) noisy, 3) volume clarity, and 4) sound effects. Furthermore, the analysis description from the speaker aspect is as follows.

1) Clarity

Audio media consists of three main elements, namely the elements of words, music, and sound effect (Sugiyono, 2009). In the UKBI test, the recorded sound in the audio was good and loud, and there was no disturbing noise. Likewise, the Learning Evaluation Test recording was in accordance with the topic and material. There was an introduction, greeting, and vocabulary that was easy to understand for BIPA level 1, or a noisy recorded sound so that exam participants could focus on the information conveyed from the listening dialogue. Both the UKBI and TEB tests had good and clear-recorded audio quality, so the information conveyed was clear.

2) Noisy

There was no noise in the listening session in the UKBI test and Learning Evaluation Test for Foreign Speakers Level 1. Both tests had good audio quality.

3) Volume clarity

In UKBI, there were eight simulations, all of which used human voices. Overall, its delivery was good by paying attention to various elements such as appropriate intonation, clarity of loud volume, and sufficiently clear sound, even though it used speakers' outputs when listening to it. In addition, the emphasis on each word related to the topic presented was easy to understand.

In the Indonesian Language Learning Evaluation Test for Foreign Speakers Level 1,

the quality of the voice recording was also good. There was no intermittent dialogue sound. There were sixteen simulations, all of which used human voices with dialogue conducted between women and men. Overall, the delivery was good by paying attention to elements such as appropriate intonation, loud volume, and clearly heard.

The quality of the recording in the form of volume clarity on the UKBI test and Learning Evaluation Test was good because the audio was loud with clarity of pronunciation, intonation, and appropriate pause emphasis.

4) Sound effects

The recording on UKBI had good quality with clear sound. There were additional sound effects. According to Sungkono (2015), sound effect is the sound other than words and music. The atmosphere in each listening model was simulated to make listeners feel the real atmosphere reflected in the audio. Additional sound effects existed on UKBI, namely the additional voices of people talking, vehicle sounds, crying babies, and laughing babies.

Audio media consists of three main elements, namely the elements of words, music, and sound effects (Sugiyono, 2009). There were some listening sessions on the UKBI using music as an identifier for an event and giving color to an event (Sugiyono, 2009). In the second and third monologues, audio music had additional music audio to give a colorful situation. The additional music was in the form of music for dances (monologue of Molulo Dance) and background music, such as the delivery of news information in a program or in an event/activity, such as audio playback at a museum (monologue about aircraft).

Like the UKBI test, the Learning Evaluation Test had good sound recording quality. There was no intermittent dialogue. In Indonesian Language Learning Evaluation Test for Foreign Speakers Level 1, sixteen simulations were using human voices with dialogue conducted by women and men. Overall, the delivery was good by paying attention to elements such as appropriate intonation, loud volume, and clearly heard.

Audio listening in the Indonesian Language Learning Evaluation Test for Foreign Speakers Level 1 only used music as a marker for the start and end of the recording. Meanwhile, dialogue listening contained in the audio did not use tones or music to add audio to the listening background so that the examinees could focus only on the dialogue contained in the listen.

The Profile of Standardized Listening Competence Evaluation Tool in Foreign Language

The profile description of listening competence for foreign languages referred to here is a listening evaluation tool made by some institutions of language evaluation tool development in their respective countries, such as America, Asia, and Europe. In the analysis, several things related to the profile of the competence evaluation tool for listening to foreign languages in several countries were discovered representing three continents: Asia-Africa by analyzing the listening competence evaluation tool from JLPT (Japanese Language Proficiency Test), America by analyzing the listening competence evaluation tool from TOEIC (Test of English for International Communication), and Europe by analyzing DELF (*Diplôme d'Etudes en Langue Française*) listening competence evaluation tool. Those three listening competence evaluation tools produced the following findings.

Speaker (Presenter Stimulus Speaker)

The indicators analyzed in the aspect of the speaker, or presenter stimulus speaker consisted of 1) pronunciation, 2) intonation, 3) vocal, 4) expressions, and 5) accuracy of pause/idea unit with the following description:

1) Pronunciation

In the JLPT listening evaluation tool, the speakers pronounced the test using the standard Japanese language, namely Japanese, with a Japanese accent. However, in the TOEIC evaluation tool, the speakers pronounced the test using an American English accent. Meanwhile, in the DELF evaluation tool, the speakers pronounced the test accent using standard French, namely French from Paris (*L'accent Parisien*). Of those three foreign language evaluation tools for listening competence, it was found that the pronunciation is spoken on the test using standard Japanese with a Japanese accent. The TOEIC pronounces the test in an American English accent, and the DELF evaluation tool pronounces the test in standard French spoken by the people of Paris (*L'accent Parisien*).

2) Intonation

In the JLPT listening evaluation tool, the speaker emphasized the high or low tone in accordance with the utterance made by Japanese native speakers with Japanese accents. Pronunciation in the JLPT test depended on the level tested. On the N5 test, it was pronounced slowly, while on the N4 test, it was pronounced a bit slower. On the N3 test, it was pronounced at a normal pace. In the TOEIC listening evaluation tool, the speaker emphasized the high or low tone in accordance with the utterance performed by

native speakers of an American English accent. Besides that, in the DELF listening evaluation tool, the speaker emphasized the high or low tone in accordance with the speech performed by Paris native speakers of Standard French (*L'accent Parisien*). Pronunciation in the DELF test depended on the level tested on the slow A1 and A2 tests, moderate B1, and normal B2. The result of the three listening competence for foreign language evaluation tools was that the speaker emphasized the high or low tone in accordance with the speech conducted by Japanese native speakers with a Japanese accent on the JLPT listening evaluation tool. The emphasizing tone was marked with a high or low tone in accordance with the speech performed by native English speakers with the American English accent on the TOEIC listening evaluation tool. Besides that, pitch stress was marked as high or low in tone in accordance with the level of the DELF listening evaluation tool tested.

3) Vocal

In the JLPT listening evaluation tool, the speakers produced vocals clearly. The vocal sounds spoken were vocals as well as vowels in Japanese. For the TOEIC listening evaluation tool, the speakers produced vocals clearly. The vowel sounds uttered were vowels like the vowels in the American English accent. Whereas in the DELF listening evaluation tool, the speakers pronounced the words clearly. The spoken words were the vowel sounds and nasal sounds. The speakers also pronounced French words or phrases as in the rules for pronouncing *Les liaisons* and *l'enchaînement*. In French, many words are homonyms but can be distinguished by context. Those three foreign language evaluation tools for listening competence found that the speaker's vocals were already good, both in the JLPT evaluation tool, which had spoken the test with clear vowels as well as vowels in Japanese. Besides that, the TOEIC evaluation tool spoke the test with clear vowels as letters pronounced American English accent, and the DELF evaluation tool, spoke out vowels and nasal sounds clearly.

4) Expression

The next indicator on the part of the speaker or the delivery of the presentation stimulus in the speaker is the expression. Expressions in the JLPT evaluation tool could be seen from the utterances performed by the speakers, namely expressions in accordance with Japanese native speakers in Japan. Expressions in the TOEIC listening evaluation

tool could be seen from the utterances performed by the speakers, namely expressions in accordance with English native speakers in America. Expressions in the DELF listening evaluation tool could be seen from the utterances performed by the speaker, namely the expressions in accordance with Standard French native speakers, namely French from Paris (*L'accent Parisien*). Those three listening competence evaluation tools for foreign languages in JLPT, TOEIC, or DELF already had good speakers' expressions.

5) Accuracy of pause/idea unit

The next and last indicator on the part of the speaker or the delivery of the presentation stimulus in the speaker is the accuracy of the pause/idea unit. In the JLPT test, the speakers were very precise in giving pause for each question. Besides that, from one to another question, it was marked by sound and instruction. Therefore, test takers could understand the transfer of one to another question and from one to the next session. In the TOEIC test, the speakers were also very precise in pausing for each unit. Therefore, the test takers could understand the transfer of units from one unit to another. In the DELF test, the speakers were very precise in pausing for each unit, from one to another exercise, marked with sounds and instructions. Therefore, the test takers could understand the transfer of units from one unit to another. Those three listening competence evaluation tools for foreign languages in JLPT, TOEIC, or DELF already had very good speakers' accuracy of pause/idea. Giving pauses for each unit was included in each evaluation tool, apart from that for one another exercise marked with sounds and instructions to make it easier for test takers.

Content

The indicators analyzed in the content aspect consisted of 1) situation/register-based language variations, 2) language use, and 3) information delivery. Furthermore, the content aspects are as follows.

- 1) Situation/register-based language variations
Situation/register-based language variation is one of the indicators in the content section. Situation/register-based language variation in the JLPT N5, N4, and N3 tests was spoken based on daily communication situations. Meanwhile, the daily communication situation in the JLPT N3 test was delivered informally as JLPT N5 and N4. The language variation in the TOEIC test was spoken based on the daily communication situation of English speakers in America. Meanwhile, language

variation in the DELF A1, A2, and B1 tests was spoken based on daily communication situations. The communication situation in DELF B2 was in a professional environment. The register used in this DELF test was *Registre courant (Le langage courant)*.

Of the three listening competence evaluation tools for foreign languages, situation/register-based language variations were found. JLPT listening evaluation tool used daily situation-based speech, but for JLPT N3, the daily communication situation was delivered informally, not the same as JLPT N5 and N4. Then, the TOEIC listening evaluation tool used the daily communication situation of English speakers in America, and the DELF listening evaluation tool used the daily communication situation and professional communication situation at level B2.

2) Language use

The next indicator in the content section is language use. In the JLPT listening test, the language used in the test content is standardized Japanese language in a Japanese accent. In the TOEIC listening test, the language used in the test content is English with an American accent. Whereas in the DELF listening test, the language used in the test content is standardized French, namely French for Paris people (*L'accent Parisien*). Different language use was found among the three listening competence evaluation tools for foreign languages. For each test, the JLPT evaluation tool used standardized Japanese in a Japanese accent, the TOEIC evaluation tool used English in an American accent, and the DELF evaluation tool used standardized French by Paris people (*L'accent Parisien*).

3) Delivery of information

The next and last indicator in the content section is the delivery of information. The speakers clearly explained the delivery of information in the JLPT listening test content. The speakers clearly explained the delivery of information in the TOEIC listening test content. Then, the speakers clearly explained the delivery of information in the DELF listening test content. Of those three listening competence evaluation tools for foreign languages, the information delivery was found to convey very clear information for each listening evaluation tool: JLPT, TOEIC, and DELF.

Audio (Recording)

The indicators analyzed in the content aspect consisted of 1) clarity, 2) noise, 3) volume clarity, and

4) sound effects. Furthermore, the audio (recording) aspects are as follows:

1) Clarity

Clarity is an indicator in the audio (recording) section. On the JLPT listening test recording, every question and instruction on how to do the test was explained very clearly. On the TOEIC listening test recording, every question and its instruction was explained very clearly. Every question or instruction was explained very clearly on the DELF listening test recording. Of the three listening competence evaluation tools for foreign languages, the clarity of audio recording was found very clearly for each listening competence evaluation tool: JLPT, TOEIC, and DELF.

2) Noisy

The next indicator in the content section is noisy. There was no noise in the JLPT listening test audio recording. There was also no noise in the TOEIC listening test audio recording. Besides that, there was no noise in the DELF listening test audio recording. Based on the result, it was found that there was no noise in JLTP, TOEIC, and DELF listening competence evaluation tools.

3) Volume clarity

The next indicator in the content section is volume clarity. The JLPT listening test audio recording has volume clarity because the speakers said it very loudly and clearly. The TOEIC Listening test audio recording also has volume clarity because the speakers said it very loudly and clearly. Besides that, the DELF listening test audio recording has volume clarity because the speakers said it very loudly and clearly. Based on the result, it was found that the volume clarity was very loud and clear in JLPT, TOEIC, or DELF listening evaluation tool.

4) Sound effects

Sound effects is the next and last indicator in the audio (recording) section. In the JLPT listening test, sound effects were presented for each question. Sound effects were given in accordance with the situation presented in the questions. This supported the recording and created a stimulus for the test takers. The TOEIC listening test also presented sound effects for each question. Sound effects were given in accordance with the situation presented in the questions. This supported the recording and created a stimulus for the test takers. Besides that, in the DELF listening test, sound effects were presented for each question. Sound effects were given in accordance with the situation

presented in the questions. This supported the recording and created a stimulus for the test takers. From those three listening competence evaluation tools for foreign languages, it was found that the sound effects were presented in the questions for JLPT, TOEIC, or DELF listening evaluation tool to support recording and create a stimulus for test takers.

CONCLUSION

In analyzing Indonesian listening competency evaluation tools such as UKBI and TEB used by foreign speakers, it is concluded that the test kits show conformity to predetermined indicators such as pronunciation, intonation, vowels, expressions, and accuracy of pause/idea units. However, in the intonation section, the pause given was too fast from one to another dialogue, and in the narrator section.

Clear audio characteristics, no noise, and clear volume appeared in the test kits of the listening competence evaluation tool. However, one of the test kits did not use any sound effects on the recording, so the test takers were less stimulated. Besides that, the relationship between time in the continuity of the listening test included the overall duration, each question duration, all question duration, and answer duration.

The competencies presented in this listening test were (1) recognizing simple words and phrases related to oneself, environment, and daily activities; (2) responding, identifying, and detailing ideas from expressions that were often used in public places; (3) understanding information coming from discussions, speeches, news, electronic media, and short films; (4) understanding explanatory texts related to social, academic, and professional spheres, and (5) understanding data and speech accompanied by language accents from various presentation.

Of those two listening tests, one of them did not use daily speech and raised more scientific topics. One of the two tests did not use the background sound even though it related to daily conversation. The delivery of information in a dialogue or a monologue was quite clear; however, the question information was not read out, so the test takers had to look for it themselves because it was not structured.

The clear audio characteristics, not noisy, volume clarity, and sound effects were clear and appropriate to the presented situation so that they could support the recording and create a stimulus for the test takers. There was an additional background sound to each dialog in the test.

The suitability of listening content included language variations in accordance with the speech situation and theme. The presented competencies in this listening test were (1) to understand words and expressions related to oneself and the environment slowly and clearly; (2) to understand words or phrases with a high frequency related to certain fields; (3) to

understand common ideas at work, school, and also from radio and television programs, (4) to understand speeches and lectures on familiar topics, and also understand films accompanied by standard dialects, and (5) to understand all kinds of spoken language delivered with a fast tempo and standard accent appropriate language.

From the analysis, it can be concluded that four of five audio listening tests used additional sound effects. This created a stimulus for the listeners/test takers. The audio quality was seen from the clarity and loudness of the narrator in performing speech acts matching his country's accent. Audio accompanied by a background sound was important to listen to because it was able to liven up the state and situation of the language being conveyed. In addition, it can also be used to determine the location of the conversation. The lack of pauses and unstable intonation on the tune made it necessary to provide pauses and tone adjustments on the listening evaluation tool as needed.

The result of this study can be used as a basis for developing the design of an international standard BIPA listening evaluation tool in accordance with the listening competence stated in Permendikbud Number 27 Year 2017.

REFERENCES

- Al-Rawafi, A. K, Sudana, D., Lukmana, I., & Syihabuddin, S. (2021). Students' apologizing in Arabic and English: An interlanguage pragmatic case study at an Islamic boarding school in Indonesia. *Indonesian Journal of Applied Linguistics*, 10(3), 589-602
<https://doi.org/10.17509/ijal.v10i3.31740>
- Muslim, A. B., Hamied, F. A., & Sukyadi, D. (2020). Integrative and instrumental but low investment: The English learning motivation of Indonesian senior high school students. *Indonesian Journal of Applied Linguistics*, 9(3), 493-507.
<https://doi.org/10.17509/ijal.v9i3.23199>.
- Alwi, H. (2000). *Tata bahasa baku Bahasa Indonesia*. Balai Pustaka.
- Brumfit, C. (1987). *The communicative approach to language teaching*. Oxford University Press.
- Castro (2015). *Filipino, Indonesian, and Thai listening test errors*. Univ of Pittsburgh.
- Chaer, A. (2015). *Morfologi Bahasa Indonesia*. Rineka Cipta
- Chaer, A., & Agustina, L. (2004). *Sosiolinguistik*. Rineka Cipta.
- Effendy, O. U. (2006). *Ilmu komunikasi; Teori dan praktek*. PT Remaja Rosdakarya.
- Fulcher, G. (2014). *Testing second language speaking*. Routledge.
- Hubackova, S. (2012). *Listening in foreign language teaching process*. University of Hradec Kralove Czech Republic.
- Ministry of Education and Culture (2017). *Regulation of the Ministry of Education and Culture Number 27 of 2017*. Government of the Republic of Indonesia.
- Muliastuti, L. (2017). *Bahasa Indonesia bagi penutur asing: Acuan teori dan pendekatan pengajaran*. Yayasan Pustaka Obor.
- Mulyati. (2006). *Pengembangan tes kompetensi BIPA tingkat dasar*. Jurnal Universitas Pendidikan Indonesia.
- Nurkencana, W., & Sumartana, P.P.N. (1983). *Evaluasi pendidikan*. Usaha Nasional.
- Rosalina, S. (2016). *Pengembangan tes kemampuan berbahasa Indonesia ragam bisnis bagi penutur asing berbasis pendekatan integratif*. Universitas Pendidikan Indonesia.
- Sugiyono. (2009). *Metode penelitian kuantitatif, kualitatif, dan R&D*. CV Alfabeta.
- Sungkono. (2015). Pengembangan Instrumen evaluasi media pembelajaran. *Majalah Ilmiah Pembelajaran*, (2).
- Tsamaratul, J. (2011). *Pengembangan tes kompetensi membaca sebagai upaya penyiapan alat uji kemahiran Bahasa Indonesia*. Universitas Pendidikan Indonesia.
- Vandergrift, L. (2007). *Recent developments in second and foreign language listening comprehension research*. Cambridge University Press.
- Widia. (2020). *American listening foreign languages evaluation for development study of BIPA listening evaluation*.
<https://eudl.eu/proceedings/KEBIPAAN/2019>
- Yeyen. (2017). Strategi komunikasi pemelajar Bahasa Indonesia bagi penutur asing asal Filipina. *Jurnal Kajian Bahasa*, 6(2).