

Development and Assessment of an Interculture-based Instrument Model in the Teaching of Speaking Skills

Aprianoto¹, Haerazi^{2,*}

¹English Language Education, Mandalika University of Education, Indonesia

²English Language Education, Fakultas Pendidikan Bahasa dan Seni (FPBS), Mandalika University of Education, Indonesia

Received October 17, 2019; Revised November 19, 2019; Accepted November 26, 2019;

Copyright©2019 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract This study was aimed at describing the development of the instrument of the intercultural-based English speaking model for students in higher education. It was viewed from the validity, reliability, and difficulty index. This study used procedural development through the 4-D model that consists of four phases; defining, designing, developing, and disseminating. The study was prepared with 25 opened-ended questions and indicators. Thus, it was submitted to two group experts (ELT and language assessment experts) to validate the instrument aiming to meet the content validity and inter-rater reliability. The questionnaire was employed. The next stage was administered an empirical test to various private universities having an English department in West Nusa Tenggara, Indonesia by stratified purposive sampling. The data were analyzed by applying content validity formula, Agreement Kappa, Person's correlation, Cronbach's alpha, and difficulty index. The result of content validity was 0.695, Cronbach's alpha was 0.756, Person correlation was around 0.40 into 0.60, and inter-rater reliability was 1.00. In addition, the difficulty index of the instrument items was around 20 into 90. It can be concluded that the instrument of intercultural-based English speaking assessment was able to be applied in the teaching of speaking.

Keywords Speaking Assessment Test, Intercultural Dimension, Validity, Reliability

1. Introduction

Teaching speaking skills is carried out discretely in Indonesian higher education including at the English language education study program. It is also taught gradually from 1st semester to 4th semester. The students cannot continue from the next stage if they do not pass from the basic stage, from the basic stage to the next one,

etc. The last stage for students attending the speaking subject is the advance level in the 4th semester. It indicates that teaching speaking skills need four semesters for students to acquire communicative competences. This learning structure is also enacted for the other three language skills, namely listening, reading, and writing skills at Indonesian higher education.

To decide that the students can continue to the next stage, the lecturer should conduct a speaking test. The assessed aspects commonly for speaking classes are addressed to fluency, accuracy, pronunciation, grammar, vocabulary, and gesture [1]. It was adopted by some researchers in assessing speaking skills in Indonesia and event in the world. For instance, Leaper and Brawn [2] detect the development of speaking proficiency with group oral test viewed from indices of complexity, accuracy, and fluency. In addition, Yan, Cheng, and Ginther [3] examine the impact of task type to investigate speaking proficiency including fluency, accuracy, and pronunciation. Isbell and Winke [4] examine students' oral proficiency in relation to the rate of speech (fluency), lexical diversity (vocabulary), and shorter silent pauses. The same thing done by [5], they explore the issue by examining lexical and discursive features for Spanish performance in oral proficiency.

Assessing the four language skill is influenced by the development of language teaching and learning nowadays. ELT Experts have developed various instructional models followed by providing the way the language skills are measured. Isbell, Winke, and Gass [6] provide face-to-face oral proficiency interview (OPI) assessment model for speaking proficiency and also as stated in [4] develop the ACTFL OPI computer model to assess students' oral proficiency and monitor students' progress in a tertiary foreign languages program. Yoo et al. [7] apply the score equity assessment (SEA) to evaluate fairness in English speaking for test-takers from a diverse cultural, linguistic, and educational background. In addition, Jensen, Grajeda, and Haertel [8] develop classroom assessment of

sociocultural interaction (CASI) model to measure cultural dimensions of classroom interaction. With respect to these mentioned models, the language assessment carried out by researchers and teachers indicates how language skills are taught.

Intercultural language learning in assessing language skills should meet learning processes provided for students both the assessed aspects and the instructional materials given. Liddicoat [9] and Haerazi [10] developed intercultural language learning for ELT learners, but the assessment model was still focused on what Brown suggested [1]. The intercultural language learning model could have specific aspects that are evaluated such as intercultural dimensions as attitude, knowledge, skills of interpretation, skills of interaction and relation, and critical awareness [11]. It is in line with what Jensen et al. [8] state the effective instruction in the teaching of TEFL classroom interaction should pay more attention to cultural dimensions. Therefore, this study was focused on developing and evaluating the instrument of assessing speaking skills within the intercultural language learning model in higher education.

The learners' intercultural communication competences are seen when learners perform their communicative abilities reflecting the five intercultural dimensions. These competencies, according to Byram [12], are acquired by implementing the intercultural language learning model. Liddicoat [13] recommends four teaching stages of the model, which are noticing, comparing, reflecting, and interacting stage. Based on the Liddicoat's model, Haerazi [10] develops the model in teaching writing skills into six teaching stages which are noticing, comparing, reflecting, concluding tentatively, constructing, and presenting. For speaking skills, this study uses Liddicoat's model with conducting modification in practice.

Implementing the intercultural language learning model in speaking classes should be followed by the appropriate speaking assessment. Some researchers conducted various researches dealing with the intercultural instructions. Reese et al. [14] and Roorda et al. [15] provide their students with cultural dimensions of Spanish native language in K-2 Latino-majority in rural California to improve classroom interaction. Equitable teaching for cultural dimensions in the classroom was conducted by Jensen [16], Jensen et al. [8], and Jensen, Meijia-Arauz, & Zepeda [17]. They claim the development of instruction was influenced by generic and cultural dimensions of classroom interactions. Even so, there is nothing an instructional assessment model focused on intercultural activities dealing with intercultural dimensions.

The intercultural dimensions proposed by some experts are elaborated as indicators of the developed instrument items for assessing speaking skills. Intercultural elucidation needs to be conducted in any aspect of English teaching and learning. As stated in [18] that intercultural elucidation in a multicultural teaching and learning context

can truly attain a true mutual understanding among learners.

This study was addressed to develop the instructional assessment instrument for intercultural-based speaking teaching. At the end of this research, the final product of this study is an instrument that consists of indicators to be utilized for the speaking skill assessment. The novelty of this study lies in the utilization of cultural dimensions combination between students' own culture and the target cultures, which are reflected in the instrument of speaking proficiency. More specifically, this study is to generate and validate the instrument of the intercultural language learning model in private universities in West Nusa Tenggara, Indonesia.

2. Literature Review

2.1. Developing Teaching Speaking Skills in Higher Education

As theory and research in intercultural language learning advance, debate continues about how to measure the intercultural language learning model. Learners can be stated they have good speaking skills if they can perform their speaking in good fluency, pronunciation, and accuracy [19]. However, the relevant studies investigating learners' speaking development and assessment are limited [20]. In addition, teachers are inquired to find a way and instrument to assess speaking skills reflecting the intercultural competences. In Indonesian higher education, teaching the four language skills is taught in the discrete instruction. The four language skills are allotted in a different semester in the English language department both in state universities and private universities in Indonesia. Consequently, lecturers should meet an appropriate way to assess the four skills. Therefore, the instrument of evaluating those must be designed and developed in line with the teaching and learning purposes. In addition, the model of language proficiency has been influential in language testing in terms of the four language skills [21], [22]. It is in line with what Isbell, Winke, and Gass [6] state language tests are useful for many purposes, including monitoring language instructional progress. In doing so, ACTFL OPIc (American Council on the Teaching of Foreign Language Oral Proficiency Interview Computer) is applied. The interview of the assisted computer is used to find out the speaking proficiency.

2.2. Speaking Test

Some studies address the language assessment through various ways. Jensen et al. [16] and Jensen, Grajeda, and Haertel [8] organize a speaking proficiency test into three domains; life application, self in a group, and agency. The first domain addresses how the interaction in the classroom

explores and values the aspects of learners' out-of-school lives to create a personal connection with classroom content [23]; & [24]. The interactions are scored from disconnected to well-connected. Thus, the associated dimensions valued include language use, difference appreciation, equity, and content personalization. Meanwhile, the domain of self in group and agency addresses how the classroom leads students to relate to and work with one another. It aims to motivate learning and to foster social identities. The research findings presented the speaking proficiencies were influenced significantly by those domains in classroom interactions.

The use of score equity assessment (SEA) is also applied in ELT classes to evaluate language proficiency. For instance, Yoo, et al. [7] utilized the TOEIC test as the assessment instrument for assessing English language skills in the workplace and also the selected-response item format in TOEFL test is applied to meet students' reading comprehension [25]. The use of TOEFL writing tasks and actual academic tasks is employed to investigate students' performance [26]. The aspects of assessment are addressed to the quality of the first and final draft viewed from the dimension of grammatical, cohesive, rhetorical, socio-pragmatic, and content control. In addition, response item formats are employed in investigating students' reading and speaking comprehension. Even so, because these processes operate simultaneously at various stages in the intercultural language learning model, operationalizing the construct and reliability of speaking proficiency is challenging for assessment purposes. This study lies in the use of intercultural dimensions between students' own cultures and the target cultures to generate the indicators for speaking assessment in higher education.

2.3. Intercultural Language Learning Assessment

One of meaningful achievement of internalization tasks at higher education is to develop learners' intercultural competences. This study tries to meet an appropriate assessment reflecting cultural competences. Some studies are conducted to find out the assessment of intercultural competence as a learner outcome of internationalization in higher education. For instance, Rogers [27] carried out an assessment of intercultural competence using program logic model that consists of four steps, namely "inputs, activities, outputs, and outputs of internationalization". It is different from Deardorff [28] who proposed five components of it, namely "world knowledge, foreign language proficiency, cultural empathy, approval of foreign people and cultures, and ability to practice in international setting".

Pre-service teachers whether domestic or international need to be treated with increasingly complex knowledge and skills to increase mobility and multiculturalism. It is important learning in higher education to provide them with opportunities to develop and practice intercultural

competence [29] [30]. As stated in [31], the intercultural competence to interact with people from different cultures is not only desirable but a necessity culturally. To reach this competence, EFL lecturers should apply an appropriate instructional model, namely the intercultural language learning model.

The intercultural language learning model is increasingly applied in universities that have an English language program including private and public Indonesian universities. For them, intercultural understanding is recognized as core learning. Because of this, the intercultural dimensions should be integrated into language teaching [32]. Additionally, global knowledge is demanded to interact among nations and peoples to communicate through online communication [33]. Based on these demands, intercultural learning is a solution for English learners to acquire those dimensions. Garson [34] defines intercultural learning is an instruction that guides the development of intercultural competence and awareness of learners' own and other cultural preferences.

3. Research Method

3.1. Research Design

This study was procedural development. It is aimed at producing a final product through some steps that should be followed. The procedural development of this study is a descriptive study using the 4-D model [35]. The final product of this study is an intercultural-based assessment model. The procedural development to attain the final product covers the stage of defining, designing, developing, and disseminating.

3.1.1. Defining

The step of defining is carried out in some research activities. First, researchers analyze the lecturers' speaking characteristics that suit the assessment model to detect the intercultural dimensions for the fourth semester of English language education programs in private universities. Afterward, the researchers come to the prior knowledge of students about the target cultures and their own cultures. It needs to consider because it becomes a minimum competence that should be acquired. So the assessment arrangement is going to be easier to decide the intercultural communication indicators that need to be attained. Second, the researchers review the related literatures to develop the intercultural-based assessment model based on the teachers' and students' speaking characteristics at the English language department in private universities. In doing so, this study is assisted by four experts, two instructional experts, and two evaluation experts.

Those experts are asked to evaluate the targeted needs and the learning needs in intercultural language instruction. The results are then utilized as a consideration to form the

assessment model for speaking skills. Third, the researchers determine the purpose of the assessment development to evaluate the intercultural dimensions in the teaching of speaking skills in the form of an open-ended question.

3.1.2. Designing

The second step is designing questions dealing with intercultural competences for speaking skills. These activities are carried out in some steps. First, the researchers conduct a deep analysis of the English department curriculum for 4th-semester students in private universities. These activities determine the instructional goals, learning objectives, and competency indicators. Second, the researchers describe the answer criteria are used to enact the intercultural communication competence or not, designing the introductory questions, question scripts, and speaking assessment rubrics. Third, the research teams write questions.

3.1.3. Developing

The third step is developing the assessment model for speaking skills. The assessment model is validated through expert judgments. Experts are invited to see the content validity of the items. The content validity of the items of speaking assessment is administered by providing the experts on ELT methodology and evaluation experts from higher education. The involved experts are based on the experience in the TEFL field and their research experiences in language testing. Those include 4 experts that consist of two ELT professors and two evaluation Ph. D lecturers. Those experts are asked to review the truth of each item of questions and assessment rubrics.

Each item is seen from substance to learning objectives and competency indicators and cultural aspects both the target culture and students' own cultures; and linguistic aspects in each question. The experts are asked to review each item of those and give comments. The suggestions and comments of those experts become considerations to complete the assessment aspects of each item. Afterward, the assessment instrument is going to be tested.

3.1.4. Disseminating

To find out the effectiveness of the assessment model, the researcher conducted dissemination to English speaking lecturers in other state universities and private universities in West Nusa Tenggara, Indonesia. In doing so, the team researcher printed the final product of this study to share in various workshops dealing with language testing. In addition, the dissemination process involved English pre-service teachers coming from public and private universities.

3.2. Research Instrument

The products of this study cover the item construction and item format for assessing speaking skills reflecting the intercultural competences. Because of it, the questionnaire

was employed to find the appropriate item construction and item format. The questionnaire was distributed into the language evaluation experts and the ELT practitioners. The description of the two product of this study can be elaborated as follows.

3.2.1. Item Construction

The items of speaking assessment questions are developed in line with the intercultural dimensions, learning objects, and content standard of ELT curriculum in private universities for the 4th-semester students. Five intercultural dimensions are elaborated into 3-4 competence indicators. Those are knowledge, attitude, skills of interpreting and relating, skills of discovery and interaction, and critical cultural awareness. Meanwhile, the indicators of those consist of 14 items. The item construction of the instrument derives from intercultural dimensions without avoiding the aspects of the complexity, accuracy, speed fluency, breakdown fluency, and repair fluency in speaking. The instrument is tailored made for developing pre-service teachers' oral proficiency in terms of the intercultural speaking assessment model.

3.2.2. Item Format

The intercultural language learning assessment consisted of 20 questions for speaking skills. Students are said that they have intercultural communication knowledge if they perform the indicators of intercultural competences. Additionally, the dynamic assessment was employed to form the items. It was in line with what Estaji and Farahanynia [36] found that the dynamic assessment has a meaningful contribution to creating the assessment of culturally diverse student-teachers. It was provided in order to anticipate student-teachers who have less willing to communicate. Therefore, the item format is designed to create interaction among student-teachers and lecturers in the speaking class.

The intercultural dimensions were developed into indicators reflecting the intercultural communication. The intercultural insertion is seen from student-teachers' attitude or knowledge when they interact and communicate with their colleagues who have different cultures. The communication skills were crucial parts of intercultural competence [37]. Therefore, this part was also considered in the item format of the developed instrument. For instance, the intercultural dimension of knowledge was broken down into 3 indicators. It can be seen in Table 1 as follows.

Table 1. Intercultural dimension of "knowledge" indicators

Intercultural Dimensions	Competence Indicators
Knowledge	Learners know when to speak, when not, what to speak about, what not, with whom, and in what manner.
	Learners have a perception of how other people perceive them in communication.
	Learners know the cultural products and practices in their own and other cultures.

3.3. Technique of Analyzing Data

This study developed an intercultural-based speaking assessment which is in the form of competence indicators. The instrument of competence indicators was valid and reliable. The validity of the test was carried out by seeing the content validity index and validity between items. Gregory [38] states the content validity of a test can be seen from the expert agreement index. The calculation was done by considering two categories that are very relevant and less relevant. The calculation of the expert agreement index was a comparison between the numbers of items to all of the item numbers (very relevant and less relevant). The items come from very relevant by two experts. The index was around 0 to 1. The criteria of the test have good content validity because the index minimum was 0.50.

The validity of each item was valid because the correlation score is higher than the sig. level 0.05. Besides, the reliability of the test was reliable because the coefficient reliability (0.70) was higher than the sig. level 0.05. It is in line with What Ozdamar [39] states the instrument is accepted as high coefficient reliability if the score of coefficient reliability is higher than 0.50. The analysis was done using the SPSS 21 device.

4. Research Findings

4.1. Content Validity

The development of an intercultural-based speaking assessment was conducted in the form of a criteria test that was validated by ELT experts and ELT assessment experts. Those experts investigated the compatibility between competence indicators which are developed by the English department in private universities. In addition, the experts are asked to validate compatibility between intercultural dimensions and competency indicators. Four experts are involved to validate the competence indicators for the intercultural assessment model. Based on experts' validation, the thirteen competence indicators were relevant for teaching speaking and assessment in the

intercultural language learning model. The result is drawn in Table 2 as follows.

Table 2. Expert validation of indicator number data

Indicator criteria	Indicator number	Total
Relevant by the two-group expert	1, 3, 6, 7, 8, 11, 12, 13, 14, 17, 19, 22, 24, and 25	14
Very relevant by group expert 1, but less relevant by group expert 2.	-	0
Very relevant by group expert 2, but less relevant by group expert 1.	-	0
Less relevant by the two-group expert.	2, 4, 5, 9, 10, 15, 16, 18, 20, 21, and 23	

In this study, the test of the content validity index was obtained 0.70. Thus, it was continued to an empirical test. The speaking lecturers are asked to answer the questions for 1.47 minutes. The highest score was 5 and the lowest score was 1 for each item. The score was calculated using bivariate value among the scores. In doing so, the SPSS 2.1 program was employed. Based on the result of the analysis, 14 indicators were valid. It means that the indicators are able to measure something that needs to be assessed. Hence, 11 indicators were invalid. The invalid one was measured based on the minimum correlation value that is required. Additionally, there were some suggestions given by the experts as the tester should determine the best strategy to assess intercultural competence is through a mix of qualitative and quantitative measures.

4.2. Reliability

The reliable test was done using the inter-rater assessment. The inter-rater reliability was obtained at 1.00. It means that the indicator test was reliable according to the two group experts (ELT experts and Language assessment experts). In other words, the intercultural competence indicators of the speaking instrument were applicable for assessing speaking skills. It was analyzed using SPSS IBM 2.1. The result can be seen in Table 3 as follows.

Table 3. Inter-rater reliability

Symmetric Measure			Asymptotic Standardized Error	Approximate Tb	Approximate Sig.
	Value				
Measure of agreement	Kappa	1.000	0.000	5.856	0.000
N of Valid Cases		25			

After the inter-rater reliability was found, this study was continued to analyze the coefficient consistency. It was analyzed using Cronbach's alpha. Based on the result of the analysis, the score of coefficient consistency was 0.756. It means that the indicator items were reliable. The consistency of the test items in approximately resulting in the same value repeatedly. In other words, the indicators of the intercultural competence for assessing speaking skills were applicable for English language students in higher education. The computation was conducted using SPSS IBM 2.1. The result of the analysis can be seen in Table 4 as follows.

Table 4. Reliability test

Statistic Reliability		
Cronbach's Alpha	Cronbach's alpha of Standardized Items	Valid Items
0.695	0.756	14

The reliability test of intercultural-based speaking indicators was stated as a valid test. It was seen from two

parts, inter-rater reliability, and coefficient consistency analysis. It can be concluded that the items of each indicator of intercultural speaking assessment were reliable. An important finding of this study was that the tester can employ the items of each indicator in various ways in practice. The point touch to assess students' intercultural competence was using the intercultural dimensions such as knowledge, the attitude, the skill of interpreting and relating, skill of discovery and interaction, and critical cultural awareness. The five critical intercultural dimensions were said reliable and consistency to assess speaking skills. The complete result of the intercultural speaking assessment can be seen in Table 5 as follows.

The intercultural dimensions consist of five aspects which are knowledge, attitude, the skill of interpreting and relating, the skill of discovery and interaction, and the critical cultural awareness. These aspects are divided into 3-4 indicators for each aspect of intercultural dimensions.

Table 5. Indicators of Intercultural Speaking Assessment

No	Intercultural Dimensions	Indicators
1	Knowledge	Learners know when to speak, when not, what to speak about, what not, with whom, and in what manner. Learners have a perception of how other people perceive them in communication. Learners know the cultural products and practices in their own and other cultures.
2	Attitude	Learners respect to cultural differences and similarities with doing polite communication. Learners have curiosity, openness, and readiness to suspend disbelief and belief about other cultures and own cultures. Learners are not to assume the right and wrong cultures.
3	Skills of Interpreting and Relating	Learners are able to identify the cultural perspectives in a new cultural context. Learners are able to explain the new cultural context that they meet. Learners are able to see how misunderstanding can arise. Learners are able to resolve cultural misunderstandings.
4	Skills of Discovery and Interaction	Learners acquire new knowledge of culture and cultural practice in real-time communication. Learners are able to build cultural knowledge into a cultural practice in communicative manners. Learners are able to integrate what they find out with what they already have. Learners are able to interact with other different cultures.
5	Critical Cultural Awareness	Learners are able to evaluate critically the cultural products and practices. Learners are able to share cultural products and practices in an appropriate way. Learners are able to change their values and behaviors when they interact with others.

5. Discussion

The development of the intercultural-based speaking assessment in this study was started from doing the validity of the instrument. The instrument was the final product to be utilized in the teaching of speaking skills in private universities. The instrument was seen its validity through content and empirical analysis based on the response data of each item. It is line with what Lissitz and Samuelsen [40] state an instrument can be identified as a valid instrument through content validity and empirical analysis. In addition, Desstya et al. [41] conducted a study on developing an instrument by doing empirical analysis and content analysis to find out the response of each item.

The content validity index in this study used around from 0 to 1. The coefficient showing closer to 0 meant that the agreement index according to validators on the relevance of items with the indicators was getting lower. Otherwise, the coefficient presenting closer to 1 meant that the agreement index from validators on the relevance of items with the indicators was getting higher. The process of content validity was a rational analysis for instrument items that are going to be measured. According to Retnawati [42], content validity as the rational analysis was also to determine the representation between the items and students' competence measured. In other words, the closer the agreement index of instrument items is, the higher the instrument items to suit the indicators are. Therefore, the instrument of the intercultural speaking assessment had items that cover all content domains to be assessed.

This study conducted content validity to attain the qualified speaking assessment instrument reflecting intercultural dimensions. The dimensions were allotted in each indicator. The items of the indicator then were tested. It was aimed at finding out the content validity of the instrument items. According to Cohen et al. [43] and Polit, et al. [44], content validity is a valuable process and should be done by any researchers who intend to meet the quality of the instrument. In this study, the researchers invited two-group experts (two ELT experts and two ELT assessment experts). Besides, the instrument was tested by doing empirical testing. It was also to validate the instrument to be more valid. It was in line with Cresswell [45] who states content validity can be derived from pieces of evidence of the empirical test and the result of experts' assessment.

After the validity test, this study considered the reliability of the test. In doing so, the inter-rater reliability was employed. Based on the result of the analysis, the inter-rater reliability was attained at 1.00. It meant the reliability of the instrument includes a good category. It was proved by the result of Cronbach's alpha coefficient consistency for standardizing items was 0.756. It showed that 75.6% of the consistency of the test would be yielding the same result repeatedly. It can be concluded that the test items were reliable. It was supported by some research

findings such as stated in [41] that found the coefficient consistency was 0.902 and this meant the instrument test was a special category. In addition, Ceniza and Cereno [46], Ozdamar [41], and Gelisli et al. [47] found that the reliability coefficient consistency was more than 0.70. It showed the acceptability of the coefficient consistency was high. It was the same clime with what Thaneerananon [48] state that the instrument reliability was higher than 0.70 had a meaning that the instrument owed sufficiency to be used. In this study, the instrument test can be concluded that the test indicators had very high reliability.

The test was categorized as a reliable and valid test in this study. Afterward, it was then continued to find the difficulty index. According to Khoshaim and Rashid [49], the difficulty index is a percentage of data based on the correct answer of students. It had a strong effect on the variability test score and the test accuracy in which the score distinguished among items [50]. Besides, the lecturers' answer as the response analysis toward the indicator items was used to meet the quality of each item. Based on the analysis, the difficulty index was applied in this study from 20% to 90%. It referred to Quaigrain and Arhin [51] criteria in which the items are said a good when the difficulty index is around 40% to 60%. When the index is less than 20% (very difficult) and more than 90% (very easy), it is not firmly established criteria and needs to be modified and fixed. In this study, 14 indicator items were reflecting the intercultural language learning. There were 6 items with easy category, 4 items with medium category, and 4 items with the difficult category.

The intercultural dimensions include the knowledge, attitude, skill of interpreting and relating, skill of discovery and interaction, and critical cultural awareness. Each dimension has indicators. For instance, the skill of discovery and interaction, students are expected to be able to identify the cultural perspectives in a new cultural context. Due to the new perspective, they are able to explain it and accept it. Students having this skill of interpreting and relating are able to avoid misunderstanding in communication. It was line with Byram [12], Liddicoat & Scarina [52], and Haerazi, et al. [53] and Apprianoto, et al. [54] who state intercultural competences in communication bring students to resolve the misunderstanding among different cultural communities. In other words, students expressed their ideas to other people, culturally acceptable and linguistically accurate [8]. In this study, the researchers developed the indicators of assessing students' competence in the teaching of speaking through intercultural dimensions.

The evaluation of the intercultural language learning adapted what Kirkpatrick [55] suggested that the language evaluation consisted of four phases, namely the reaction, learning, behavior, and result phase. The indicators of the instrument are also highlighted in the evaluation phases. For example, the phase of the reaction, lecturers should

evaluate students' activities in the class whether they have the knowledge or not. From here, lecturers can see the students' reaction toward the first intercultural dimension, which is knowledge. At the end of this, students showed their reaction after getting knowledge of their own cultures and the target culture provided. Therefore, the assessment of speaking skills was seen in this process. It was in line with what Liddicoat and Scarino [52] state students' intercultural communication competence can be presented when students have prior knowledge of their own cultures and the target cultures. Due to this knowledge, they can perform communication, culturally appropriate.

The process of assessing the speaking skill in this study still keeps the aspects of micro-skills of speaking. These of speaking are maintained in the assessment process. These consist of pronunciation, fluency, accuracy, grammar, language use, and communication [56]. The aspects of those can be observed when the lecturers measure the speaking skills through intercultural language learning. For instance, when lecturers conducted the phase of evaluating the skill of discovery and interaction, students are asked to communicate the provided topic and lecturers give questions around the topic. Communication and interaction occur in this phase. At the same time, the lecturers notice the micro speaking skills performed by the students. It refers to Nurgiyantoro [22] who states the assessment model is a way to assess what lecturers' measure. In this study, the intercultural-based speaking assessment model was developed as a way to measure speaking through intercultural language learning.

6. Conclusions

This study developed the instrument of intercultural-based English speaking model for students in higher education. The instrument had indicators produced from intercultural dimensions. It was said as a valid and reliable instrument. It was supported by the result of content validity was 0.695, Cronbach's alpha was internal consistency coefficient was 0.756, Person correlation was around 0.40 into 0.60, and inter-rater reliability was 1.00. In addition, the difficulty index of the instrument items was around 20 into 90. It consisted of 14 indicators generated from five intercultural dimensions. It can be concluded that the instrument of intercultural-based English speaking assessment was able to be applied in the teaching of speaking. Besides, the phase of doing evaluation can adapt Kirkpatrick's model [55] that consisted of the phase of the reaction, learning, behavior, and result.

The researcher teams expect the instrument can be applied in the English language department. This instrument should be got along with the implementation of the intercultural language learning model in the class. This instructional model leads students to learn their own cultures and the target cultures as learning inputs. Dealing

with the steps of the model, the lecturers are recommended to adapt or adopt what Liddicoat proposed [13] and Scarino [57] applied.

Acknowledgements

This work was supported financially by the Ministry of Research and Technology of Higher Education (KEMENRISTEKDIKTI) under the research fund programs of the National Competitive Applied Research (*Penelitian Terapan Kompetitif Nasional*) with contract number; 03/K8.0576L1?SK/IKIP-Mtr/2018. The authors also thank all colleagues at IKIP Mataram for valuable comments on different parts of this manuscript.

REFERENCES

- [1] Brown, H. D. (2007). *Principles of language learning and teaching*. 5th edition. San Francisco: Pearson Longman.
- [2] Leaper, D. A., & Brawn, J. R. (2018). Detecting development of speaking proficiency with a group oral test : A quantitative analysis, *Language Testing*, 0(00) 1–26. <https://doi.org/10.1177/0265532218779626>
- [3] Yan, X., Cheng, L., & Ginther, A. (2018). Factor analysis for fairness : Examining the impact of task type and examinee L1 background on scores of an ITA speaking test, *Language Testing*, 0(00) 1-28. <https://doi.org/10.1177/0265532218775764>
- [4] Isbell, D., & Winke, P. (2019). ACTFL oral proficiency interview-computer (OPIc), *Language Testing*, 36(3) 467-477 <https://doi.org/10.1177/0265532219828253>
- [5] Brown, A. V., Cox, T. L., & Thompson, G. L. (2017). A comparative discourse analysis of Spanish past narrations from the ACTFL OPI and OPIc. *Foreign Language Annals*, 50(4), 793–807. <https://org/doi/10.1111/flan.12302>
- [6] Isbell, D. R., Winke, P. M., & Gass, S. M. (2018). Using the ACTFL OPIc to assess proficiency and monitor progress in a tertiary foreign languages program, *Language Testing*, 0(00) 1-27 <https://doi.org/10.1177/0265532218798139>
- [7] Yoo, H., & Monfils, L. F. (2018). Measuring English language proficiency across subgroups: Using score equity assessment to evaluate test fairness, *Language Testing*, 0(00) 1-21. <https://doi.org/10.1177/0265532218776040>
- [8] Jensen, B., Grajeda, S., & Haertel, E. (2018). Measuring cultural dimensions of classroom interactions. *Educational Assessment*, 23(4), 250–276. <https://doi.org/10.1080/10627197.2018.1515010>
- [9] Liddicoat, A. J. (2008). Pedagogical practice for integrating the intercultural in language teaching and learning. *Japanese Studies*, 28(3), 277–290. <https://doi.org/10.1080/10371390802446844>
- [10] Haerazi. (2018). Developing an interculture-based language learning (IBLL) model in the teaching of writing skills in

- Higher education institution. Ph. D Dissertation, Language Educational Science, Yogyakarta State University, Indonesia
- [11] Byram, M. (2012). Intercultural Competence. The Encyclopedia of Applied Linguistics. <https://doi.org/10.1002/9781405198431.wbeal0554>
- [12] Byram, M. (1997). Teaching and assessing intercultural communicative competence. Clevedon: Multilingual Matters.
- [13] Liddicoat, A. J. (2011). Language teaching and learning from an intercultural perspective. In Eli Hinkel (Ed.), Handbook of research in second language teaching and learning. 2nd edition. Mahwah, USA: Lawrence Erlbaum and Associates.
- [14] Reese, L., Jensen, B., & Ramirez, D. (2014). Emotionally supportive classroom contexts for young Latino children in rural California. The Elementary School Journal, 114(4), 501–526. <http://doi.org/10.1086/675636>.
- [15] Roorda, D. L., Koomen, H. M. Y., Spilt, J. L., & Oort, F. J. (2011). The influence of affective teacher-student relationships on students' school engagement and achievement: A meta-analytic approach. Review of Educational Research, 81(4), 493–529. <https://org.doi/10.3102/0034654311421793>
- [16] Jensen, B. (2014). Framing sociocultural interactions to design equitable learning environments. In J. L. Polman, E. A. et al. (Eds.), Learning and Becoming in Practice: The International Conference of the Learning Sciences (ICLS), Vol. 2., (pp. 903–910). Boulder, CO: International Society of the Learning Sciences.
- [17] Jensen, B., Mejía-Arauz, R., & Zepeda, A. R. (2017). Equitable teaching for returnee children in Mexico. Journal of Sinéctica, 48, 1–20.
- [18] Jie, L. (2019). Cultivation of "cultural confidence" in foreign literature course from a comparative perspective. Universal Journal of Educational Research, 7(9) 1869-1873. <https://doi.org/10.13189/ujer.2019.070904>
- [19] Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. Applied Linguistics, 30(4), 461–473.
- [20] Vercellotti, M. L. (2017). The development of complexity, accuracy, and fluency in second language performance: a longitudinal study. Applied Linguistics, 38(1), 90–111
- [21] Bachman, L. F., & Palmer, A. S. (1996). Designing and developing useful language tests. Oxford: Oxford University Press.
- [22] Nurgiyantoro, B. (2013). Evaluating the language teaching and learning. Yogyakarta: Economic Press of UGM.
- [23] Gee, J. P. (2001). Language, class, and identity: Teenagers fashioning themselves through language. Linguistics and Education, 12(2), 175–194. [http://doi.org/10.1016/S0898-5898\(00\)00045-0](http://doi.org/10.1016/S0898-5898(00)00045-0)
- [24] González, N., Moll, L., & Amanti, C. (Eds.). (2005). Funds of knowledge: Theorizing practices, households, communities, and classrooms. Mahwah, NJ: Lawrence Erlbaum Associates.
- [25] Becker, A. & Nekrasova-Beker, T. (2018). Investigating the effect of different selected-response item formats for reading comprehension. Educational Assessment, 23:4, 296-317, <https://org.doi/10.1080/10627197.2018.1517023>
- [26] Llosa, L., & Malone, M. E. (2018). Comparability of students' writing performance on TOEFL iBT and in required university writing courses. Language Testing, 0(00) 1-29. <https://doi.org/10.1177/0265532218763456>
- [27] Rogers, P. J. (2000). Program theory: Not whether programs work but how they work. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds.), Evaluation models: Viewpoints on educational and human services evaluation (2nd ed., pp. 209-232). Boston: Kluwer Academic
- [28] Deardorff, D. K. (2006). Identification and assessment of intercultural competence as a student outcome of internationalization. Journal of Studies in International Education, 10(3), 241-266. <https://doi.org/10.1177/1028315306287002>
- [29] Spitzberg, B. H., and G. Changnon. (2009). Conceptualizing Intercultural Competence." In The Sage Handbook of Intercultural Competence, edited by D. K. Deardorff, 2–52. Thousand Oaks, CA: Sage
- [30] Perry, L. B., and L. Southwell. (2011). Developing intercultural understanding and skills: models and approaches. Intercultural Education, 22 (6): 453–466. <https://org.doi/10.1080/14675986.2011.644948>
- [31] Corder, D. & U-Mackey, A. (2015). Encountering and dealing with difference: Second life and intercultural competence. Intercultural Education, 26:5, 409-424. <https://org.doi/10.1080/14675986.2015.1091213>
- [32] Ricketts, Peter, and Humphries, J. (2015). "Taking an ethical approach to Internationalization. University Affairs. <http://www.universityaffairs.ca/opinion>
- [33] Morong, G & DesBiens, D. (2016) Culturally responsive online design: learning at intercultural intersections, Intercultural Education, 27:5, 474-92, <https://org.doi/10.1080/14675986.2016.1240901>
- [34] Garson, & Dawne, K. (2013). "Are We Graduating Global Citizens? A Mixed Methods Study Investigating Students' Intercultural Development and Perceptions of Intercultural and Global learning in Academic Settings." PhD diss., Simon Fraser University. <http://summit.sfu.ca/item/14213>.
- [35] Thiagarajan, S., Semmel, D. S., & Semmel, M. I. (1974). Instructional development for training teachers of exceptional children: A sourcebook. Minneapolis: Leadership Training Institute/Special Education, University of Minnesota.
- [36] Estaji, M., & Farahanynia, M. (2019). The immediate and delayed effect of dynamic assessment approaches on EFL learners' oral narrative performance and anxiety. Educational Assessment, <https://org.doi/10.1080/10627197.2019.1578169>
- [37] Nielsen, B. L., Laursen, H. D., Reol, L. A., Jensen, H., Ana Kozina, Masa Vidmar, Maria Rasmusson, Iris Marušić, Albert Denk, Nina Roczen, Svetlana Jurko & Ales Ojstersek (2019): Social, emotional and intercultural competencies: a literature review with a particular focus on the school staff, European Journal of Teacher Education,

<https://org.doi/10.1080/02619768.2019.1604670>

- [38] Gregory, R. J. (2004). *Psychological testing: History, principles, and applications*. Needham Heights, MA, US: Allyn & Bacon.
- [39] Ozdamar, K. (2013). *Statistical data analysis by packet program (9thEd.)*. Eskisehir: Nisan Press.
- [40] Lissitz, R. W., & Samuelson, K. (2007). Further clarification regarding validity and education. *Educational Researcher*, 36(8), 482-484. <http://www.jstor.org/stable/4621104>.
- [41] Desstya, A., Prasetyo, Z. K., Suyanta, Susila, I., & Irwanto. (2019). Developing an instrument to detect science misconception of an elementary school teacher. *International Journal of Instruction*, 12(3), 201-218. <https://doi.org/10.29333/iji.2019.12313a>
- [42] Retnawati, H. (2016). Proving content validity of self-regulated learning scale: The comparison of Aiken index and expanded Gregory index, *Study and Evaluation in Education*, 2(22), 155-164. <http://journal.uny.ac.id/index.php/reid>.
- [43] Cohen, L., Manion, L., & Morisson, K. (2018). *Research methods in education (8th Edition)*. London: Routledge Taylor & Francis Group.
- [44] Polit, D. F., Beck, T., & Owen, S. V. (2007). Focus on study methods: is the CVI an acceptable indicator of content validity? (Appraisal and Recommendations), *Study in Nursing and Health*, 30, 459-467. <http://doi.org/10.1002/nur.20199>.
- [45] Creswell, J. W. (2012). *Educational study: Planning, conducting, and evaluating quantitative and qualitative study Boston (4th Ed.)*. Boston, USA: Pearson Education.
- [46] Ceniza, J. C., & Cereno, D.C. (2012). Development of mathematic diagnostic test for DORSHS. Retrieved from [http://www.docst.edu/academic/graduate_school/5\(1\)2012](http://www.docst.edu/academic/graduate_school/5(1)2012).
- [47] Gelişli, Y., Beisenbayeva, L., & Gumilyov, L. N. (2017). Scientific inquiry competency perception scale (the case of Kazak post-graduate students) reliability and validity study. *International Journal of Instruction*, 10(1), 273-288 <http://doi.org/10.12973/iji.2017.10117a>.
- [48] Thaneerananon, T. (2016). Development of a test to evaluate students' analytical thinking based on fact versus opinion differentiation. *International Journal of Instruction*, 9(2), 123-138. <http://doi.org/10.12973/iji.2016.929a>
- [49] Khoshaim, H. B., & Rashid, S. (2016). Assessment of the assessment tool: Analysis of items in a non-mcq mathematics exam. *International Journal of Instruction*, 9(1), 119-132. <http://doi.org/10.12973/iji.2016.9110a>.
- [50] Thorndike, R. M., et al. (1991). *Measurement and evaluation in psychology and education (5th Ed.)*. New York: Mac Millan.
- [51] Quagrains, K., & Arhin, A. W. (2017). Using reliability and item analysis to evaluate a teacher-developed test in the Educational Measurement and Evaluation. *Journal of Educational Assessment & Evaluation: Research Article*, 4(00), 1-11. <http://dx.doi.org/10.1080/2331186X.2017.1301013>.
- [52] Liddicoat, A. J., & Scarina, A. (2013). *Intercultural language teaching and learning*. Melden, Belgium: Wiley Blackwell.
- [53] Haerazi, Irwansyah, D., Juanda., Azis, Y. A. (2018). Incorporating intercultural competences in developing English materials for writing classes, *Journal of Language Teaching and Research*, 9(3), 540-547. <http://dx.doi.org/10.17507/jltr.0903.13>
- [54] Aprianoto, Maurisa, S., & Haerazi. (2018). Developing instructional tasks and materials of the intercultural-based English speaking skills (IBES) model, *Journal of Arts, Science, & Commerce*, Vol. IX (October), 57-64. <http://dx.doi.org/10.18843/rwjasc/v9i4/08>.
- [55] Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating training program (3rd Edition: The four levels)*. California: Baret-Koehler Publisher.
- [56] Richards, J. C. (2015). *Key issues in language teaching*. Cambridge: Cambridge University Press.
- [57] Scarino, A. (2009). Assessing intercultural capability in learning languages: Some issues and considerations. *Language Teaching*, 42(1), 67-80. <https://doi.org/10.1017/S0261444808005417>