



Developing an assessment model to improve the Sundanese speaking skill of Indonesian middle school students

Usep Kuswari*, Dingding Haerudin, Haris Santosa Nugraha

Universitas Pendidikan Indonesia

*Corresponding Author: usep.kuswari@upi.edu

ABSTRACT

The study aims to develop an assessment model of Sundanese language speaking skills for junior high school students in West Java province of Indonesia. This test model is expected to provide an effective assessment program which includes contents, scope, delivery format and time. A design-based research was adopted to produce a model of class-based Sundanese language and literature learning assessment via several tryouts, interviews, and expert judgement. Findings of this study show that the developed Sundanese Speaking Test Model (SSTM) fits the criteria of a good test in terms of clarity of instruction, clarity of test materials and assessment, use of communicative and standardized language, and use of correct varieties of spoken language. Statistical calculations of field tryout scores show that the assessment model was relevant to field data, indicating that the model is fit and worth using as an assessment instrument of Sundanese speaking skill of junior secondary school students as supported by theoretical concepts and expert judgment. Results of judgment by experts, users, and practitioners also show that the SSTM is reliable for use as an instrument for assessing students' speaking skill at middle school level.

Keywords: communicative competence; integrated assessment; speaking test; Sundanese language

Article history

Received:
19 October 2021

Revised:
26 December 2021

Accepted:
8 January 2022

Published:
15 February 2022

Citation (APA Style): Kuswari, U., Haerudin, D., Nugraha, H. S. (2022). Developing an assessment model to improve the sundanese speaking skill of Indonesian middle school students. *Cakrawala Pendidikan: Jurnal Ilmiah Pendidikan*, 41(1), 43-54. <https://doi.org/10.21831/cp.v41i1.46170>

INTRODUCTION

Assessment and evaluation are interrelated in education. The link lies in the measurement activities of students' competence conducted by means of test or non-test in accordance with their grades or levels of ability and development that result from the learning process that the students have experienced. Students' learning process of the Sundanese language can be evaluated based on the measurement and assessment of competence and language performance data. The evaluation can be performed formally or informally. This is in line with what Brown (2004:6) has said that all tests are formal evaluation, but not all evaluations are formal tests. Evaluation can be defined as a collection of information about the quality and quantity of changes that result from students' learning. Johnson and Johnson (2002: 6) state that learning evaluation has the aims of (1) determining the level of students' knowledge and skills, (2) understanding the progress towards the learning objectives in order to help prepare an instructional program, and (3) providing data to consider the final level of the students' learning.

The quality of language learning evaluation is determined by the clarity of its objectives so that it enables the evaluator to get a genuine response from the students (Mally & Pierce, 1996). The choice and use of a suitable evaluation approach is particularly useful to learn what kind of language competence and performance to be used as evaluation criteria. The management of evaluation instrument types, i.e. test and non-test, including the recording of the test results and their presentation in numbers and fidelity in percentage; and the use of test results are necessary in the development of an evaluation instrument. Besides, test administration also includes two important aspects: format and the nature of language. This aspects are also identifiable in Sundanese language, a local vernacular in West Java province of Indonesia.

Good quality assessment needs some criteria. The first aspect is integration. A good assessment integrates all language skills in a natural way. Davies (1977) confirmed that an integrated assessment presents all language skills better than a combination or single test. For instance, an oral proficiency test needs to combine listening and speaking followed by reading and writing activities. In addition, this integrated oral proficiency skill has to require natural or authentic setting. Good assessment requires authentic tasks in natural settings.

The second aspect of quality assessment is communicative competence. Some scholars argue that communicative language competence include both language competence and the ability to apply the language competence (Bachman, 1990; Morrow; 1977; Canale & Swain, 1980). The communicative language test must be carried out according to capacity (Widdowson, 1978) or communicative language skills (Bachman, 1990). The whole act of generalizing outside of the setting in communicative language test should really tests the ability to use knowledge and/or language about it. And vice versa, the ability to see how the ability (knowing about the use of a language) may be evaluated except through its manifestation in achievement. In addition, Canale dan Swain (1980) presented a clarification of the terms used that are relevant to the formation of communicative competence. These scholars believe that communicative competence includes grammatical competence (knowledge of grammatical rules), sociolinguistic competence (knowledge about rules of uses and rules of discourse) and strategic competence (knowledge about communicative strategies). This model was later modified into a forum-dimensional model that consist of linguistic, sociolinguistic, discourse and strategic competence. This model is necessary to develop in Sundanese language assessment.

Furthermore, the existence of vernacular languages in Indonesia such as Sundanese, Javanese, and Balinese is influenced by the power of Bahasa Indonesia as the national unifying language. Many local ethnicities have lost their vernaculars due to the power of this national language. Teachers' lack of knowledge in speaking test management has resulted in less effective Sundanese language learning at junior secondary schools (SMPs) in this region of Indonesia. Consequently, many junior secondary school graduates are unable to communicate effectively in Sundanese. This is problematic when they extend their study or find a decent job. Therefore, it is necessary to discover the obstacles faced by the Sundanese language teachers. This has motivated the researcher to develop a Sundanese language learning evaluation model which is rarely used as a model. This Sundanese language learning evaluation model can be developed further into a Sundanese language learning model.

Many studies have been conducted in the area of language learning assessment, including Sundanese language and culture as a local content of Indonesian school curriculum. For instance, Koswara et al. (2021) examined the assessment of Sundanese language students in writing descriptive texts. Other scholars designed Sundanese cultural literacy programs based on national curriculum of 2013 for students of elementary schools in West Java region (Kurniasih, Komariah, & Rodiah, 2014). Another study explored the mathematical activities of estimating, measuring and making patterns among Sundanese people (Muhtadi, Sukirwan, Warsito, & Prahmana, 2017). The other study explored the quality and values of character education in Sundanese language text book of the 2013 elementary school curriculum (Haerudin, Suherman, & Nugraha, 2018).

However, these studies do not specifically address the oral proficiency of Sundanese language. Meanwhile, this skill is the most salien indicator of language proficiency. Given that Sundanese is the second largest ethnicity in Indonesia, this vernacular is one important local content of national curriculum. Teachers need to have an assessment model which can support the oral proficiency of their students in Sundanese language.

Therefore, this study aims to develop a Sundanese language speaking test instrument which focuses on its content, scope, format and time delivery. This model is necessary to support students and teachers who are involved in the teaching and learning of Sundanese language at middle schools in West Java province of Indonesia. Specifically, this study aims to describe (a) the Sundanese language speaking tests that are currently used in junior secondary schools in West Bandung District, (b) the quality of integrative approach-based Sundanese language speaking test instrument according to expert judgement; (c) results of tryout; and (d) its benefits for schools and students.

METHOD

The focus of this study is the trial of speaking assessment of Sundanese language in 23 junior secondary schools (SMP) of years 7-9 and expert judgement on this Sundanese language assessment. This assessment has 23 items consisting of various topics such as greetings, public speaking, daily conversation, story telling, and asking permissions. The tryouts were conducted at 23 public and private junior secondary schools in West Bandung Regency of West Java province. Each class had about 35-40 students. In the context of Indonesian schooling curriculum, Sundanese language is a local content which is taught at primary and junior secondary schools in West Java province of Indonesia. It is similar to other local vernaculars in Indonesia such as Javanese, Batakese, and Balinese, which are included as local contents in the school curriculum.

Speaking test is not only an oral test. It can also be a performance test, which is a non-verbal test. This means that speaking test does not only evaluate the act of speaking, but also the processes that lead to speaking. Therefore, observation technique is often considered to be part of speaking test technique. An examiner needs to use this technique to observe (not just listen) how a testee speaks. This is particularly applicable in a direct oral performance testing.

To elicit students' ability to speak Sundanese, pictures can be used as stimuli. Pictures are good stimuli for speaking skills evaluation for primary school pupils. However, picture stimuli can also be used with students with higher level language skills, depending on the pictures used. According to Oller (1979: 47-8), 308-14), good pictures are pictures that draw students' interest to speak or those that make it easier to reveal their speaking skills. The tasks can be given to the students in the form of questions or story telling task.

This type of speaking test can be in the form of (a) verbal essays, where students are asked to speak for three minutes with one or more general topics, (b) oral presentations, students are expected to speak briefly on topics they have prepared beforehand, (c) free interviews, where students have conversations to develop a model that is not structured and there are no predetermined procedures, (d) controlled interviews, (e) role playing, which is one of the speaking assessment techniques that aims to play a role in the interaction might be expected in the real world. Interaction can occur between two students or usually the examiner acts as one of them.

Adopting a design-based method developed by Borg and Gall (1983), the Sundanese speaking assessment underwent the following procedures. First, surveying and collecting syllabus and lesson plans from the target school teachers to identify the existing practice of Sundanese language teaching process. Second, planning which includes defining expertise, stating the objectives that determine the procedure sequence and the possibility of small-scale tryout. Third, developing the preliminary product design – including learning material, manual book, and the Sundanese language speaking and writing test instrument. Fourth, field tryout – organizing 1 school with 3 up to 45 subjects. Interview, observation and questionnaire data were collected and analyzed. Next, revision of assessment instrument as suggested by the results of the previous field tryout. Sixth, primary field testing – measuring 1 up to 2 schools with 6 up to 70 subjects. Quantitative data of prior and post subject sequence administration are collected. The results are evaluated by putting a value on the objective sequence and comparing the data with a control group, if appropriate. Seventh, revision of operational Sundanese speaking assessment – product revision as suggested by the results of primary field testing. Next step is operational field testing. This step organized interviews with teachers in 1 to 3 schools including 9 teachers and up to 135 students. Ninth, final product revision –revision of speaking assessment items as suggested by the results of operational field testing. Finally, dissemination and implementation – Product reporting in professional conferences and journals. To get a more specific procedures, the study also combined these procedures with Kirkpatrick model (?) into four stages, namely: (1) preliminary stage; (2) design stage; (3) tryout and revision stage; and (4) implementation stage.

Quantitative data analysis with Confirmatory Factor Analysis (CFA) was used to figure out the validity of the Sundanese language speaking test instrument. The analysis was performed with the help of LISREL. CFA is used to examine the existing construct validity (Mueller, 1966: 124). The criteria used to test the appropriateness (suitability) of the data collection instrument (measurement model) and speaking test instrument can be seen on Table 1 below.

Table 1 Goodness of Fit Statistics

Statistics	"Fit" Criteria
λ^2 - <i>Chie-Square</i>	$p > 0,05$
RMSEA	$> 0,05$
GFI	$> 0,9$
AGFI	$> 0,9$
PGFI	$> 0,9$

(Adopted from Byrne, 1991; Wang & Wang, 2012)

To test the conformity of the hypothesis model of the Sundanese language speaking test instrument with the empirical data, two indicators were used, namely: 1) *P-value* $> 0,05$; and 2) *Root Mean Square Error of Approximation (RMSEA)* $< 0,05$ (see Suhuri, 2009). The quality of the Sundanese language speaking test instrument was analyzed using *LISREL (Linear Structural Relationships)* with the implementation test data. The hypothesis model empirically tested in this study is the Sundanese language speaking test instrument model. The instrument was designed with an assumption that the speaking learning process of Sundanese language significantly influences its learning outcome.

Qualitative data analysis is the analysis of validation result data received from experts (expert judgement) and users of the Sundanese language speaking and writing test instrument who have given input to improve the evaluation model and its instruments. Analysis was performed on the construct of the test instrument, namely the manual, the material, the language, the type of test instrument used, writing procedure, and scoring guidelines. Morse (1994: 67) stated that evaluation research aimed at finding information to understand a successful intervention mechanism. Fatterman (1988: 210) suggested that after discussing and supplying examples of the techniques of the approach, the researcher should discuss the reasons for combining the techniques used. In this qualitative data analysis, the quantitative data obtained from the evaluation instrument were converted to qualitative data on the scale of 5, then described and the results of the description were made as the basis for evaluating the quality of the evaluation model that was being developed (see Suhuri, 2009). The conversion of quantitative data into a scale 5 qualitative data was made using the rules modified from rules developed by Sudiyono (2003: 329 – 339). The criteria for the interpretation of the qualitative data can be seen in the following table.

Table 2 Criteria for the Interpretation of Research Results

	Interval	Category
81-100	4,01 - 5,0	Very Good
61-80	3,01 - 4,0	Good
41-60	2,01 - 3,0	Fair
21-40	1,01 - 2,00	Poor

For data analysis of the validity of Sundanese writing test instrument, Product Moment formula was used to calculate the evaluation result correlation between expert judgement and users' judgement, and to correlate the first to the second tryout.

FINDING

Based on results of semi-structured interviews, the study identified the following 23 test items in Sundanese proficiency skill. The items are described in Table 1 below.

Table 3 Types of Sundanese oral proficiency topics

No	Sundanese oral proficiency topics
1.	Greetings
2.	Giving invitation
3.	Accepting invitation
4.	Offering something
5.	Giving a compliment
6.	Having a small talk
7.	Interviewing
8.	Giving announcement
9.	Giving interruption
10.	Requesting
11.	Expressing hope
12.	Asking something
13.	Refusing something
14.	Giving argument
15.	Threatening
16.	Complaining
17.	Accussing and refusing accusation
18.	Showing agreement
19.	Showing confirmation
20.	Giving a report
21.	Telling something or story
22.	Reporting an event
23.	Giving a speech
24.	Prohibiting something

These oral proficiency skills are considered necessary to help students develop their Sundanese oral proficiency so that they can communicate properly within their community.

Furthermore, the validity test for the Sundanese speaking test instrument was conducted three times: first and second tryout, and third tryout (implementation). In each tryout, respondents were asked to give comments on the speaking test instrument. In addition to respondent, the study also sought judgments from experts via semi-structured interviews.

The results of the tryouts were analyzed with *confirmatory factor analysis* (CFA) program LISREL 8.51. Item validity was based on the factor load value (λ) of each instrument's item, whereas the measurement model conformity was analyzed based on the significance value (*P-Value*) and *Root Mean Square Error of Approximation* (RMSEA). During the implementation stage, the analysis of the measurement model conformity was based not only on the score of each instrument's item but also analyzed with composite score. The composite score was obtained from the sum of indicator score and the sub-aspect of each speaking and writing test aspect of the Sundanese language learning.

Results of First Tryout

Conformity of hypothetical model and empirical data test was administered in reference to three indicators, namely: 1) Factor Load Value (λ), (2) Significance Probability, and (3) Root Mean Square Error of Approximation (RMSEA). Based on the data obtained from implementation test in a number of classes analyzed with LISREL 8.51, the following results were obtained:

Table 4 Results of Validity Test of Speaking Test Instrument

Results of First Tryout

Instru- ment	Load Factor Value (λ)	P-Value	RMSE A	df	Chie- Square	Validity
1	2	3	4	5	6	6
1	(λ)> 0,30	0,965	0,00	1.03	0,029	valid
2	(λ)> 0,32	0,876	0,00	1.03	0,028	valid
3	(λ)> 0,40	0,765	0,00	1.03	0,019	valid
4	(λ)> 0,25	0,665	0,00	1.03	0,021	valid
5	(λ)> 0,34	0,865	0,00	1.03	0,023	valid
6	(λ)> 0,33	0,565	0,00	1.03	0,019	valid
7	(λ)> 0,33	0,755	0,00	1.03	0,022	valid
8	(λ)> 0,43	0,565	0,00	1.03	0,019	valid
9	(λ)> 0,30	0,965	0,00	1.03	0,029	valid
10	(λ)> 0,30	0,965	0,00	1.03	0,029	valid
11	(λ)> 0,34	0,865	0,00	1.03	0,023	valid
12	(λ)> 0,30	0,965	0,00	1.03	0,029	valid
13	(λ)> 0,42	0,876	0,00	1.03	0,028	valid
14	(λ)> 0,04	0,765	0,00	1.03	0,019	valid
15	(λ)> 0,35	0,665	0,00	1.03	0,021	valid
16	(λ)> 0,34	0,865	0,00	1.03	0,023	valid
17	(λ)> 0,33	0,565	0,00	1.03	0,019	valid
18	(λ)> 0,33	0,755	0,00	1.03	0,022	valid
19	(λ)> 0,23	0,565	0,00	1.03	0,019	valid
20	(λ)> 0,30	0,965	0,00	1.03	0,029	valid
21	(λ)> 0,30	0,965	0,00	1.03	0,029	valid
22	(λ)> 0,34	0,865	0,00	1.03	0,023	valid
23	(λ)> 0,3	0,965	0,00	1.03	0,029	valid

The above table shows that the validity of the 23 speaking test items administered at various junior secondary schools (SMPs) in West Bandung District was considered good with *Goodness of Fit Statistics* criteria as follows:

Statistics	"Fit" Criteria
λ^2 - <i>Chie-Square</i>	p>0,05
<i>P-Value</i>	0,00
<i>RMSEA</i>	>0,05

(Adopted from Byrne, 1991; Wang & Wang, 2012)

Based on the tryout results, it was shown that all load factor (λ) >0,3, *P-Value* was 0,00, and *RMSEA* >0,05. This result indicates that there is conformity between the speaking test instrument and the competence being evaluated.

Results of Second Tryout

In reference to the statistical test results, the validity of speaking test instruments in the second

tryout can be seen in the following Table 5.

Table 5 Results of Vailidity Test of Speaking Test Instrument Results of the Second Tryout

Instrument	Load Factor Value (λ)	P-Value	RMSE A	df	Chie-Square	Validity
1	2	3	4	5	6	6
1	(λ)> 0,4	0,965	0,00	1.03	0,029	Valid
2	(λ)> 0,3	0,876	0,00	1.03	0,028	Valid
3	(λ)> 0,3	0,765	0,00	1.03	0,019	Valid
4	(λ)> 0,4	0,665	0,00	1.03	0,021	Valid
5	(λ)> 0,24	0,865	0,00	1.03	0,023	Valid
6	(λ)> 0,3	0,565	0,00	1.03	0,019	Valid
7	(λ)> 0,4	0,755	0,00	1.03	0,022	Valid
8	(λ)> 0,3	0,565	0,00	1.03	0,019	Valid
9	(λ)> 0,4	0,965	0,00	1.03	0,029	Valid
10	(λ)> 0,5	0,965	0,00	1.03	0,029	Valid
11	(λ)> 0,30	0,865	0,00	1.03	0,023	Valid
12	(λ)> 0,3	0,965	0,00	1.03	0,029	Valid
13	(λ)> 0,35	0,876	0,00	1.03	0,028	Valid
14	(λ)> 0,34	0,765	0,00	1.03	0,019	Valid
15	(λ)> 0,31	0,665	0,00	1.03	0,021	Valid
16	(λ)> 0,48	0,865	0,00	1.03	0,023	Valid
17	(λ)> 0,49	0,565	0,00	1.03	0,019	Valid
18	(λ)> 0,30	0,755	0,00	1.03	0,022	Valid
19	(λ)> 0,31	0,565	0,00	1.03	0,019	Valid
20	(λ)> 0,30	0,965	0,00	1.03	0,029	Valid
21	(λ)> 0,34	0,965	0,00	1.03	0,029	Valid
22	(λ)> 0,48	0,865	0,00	1.03	0,023	Valid
23	(λ)> 0,37	0,965	0,00	1.03	0,029	Valid

The above table shows that the validity of the 23 speaking test items administered at SMPN in West Bandung District was considered good with *Goodness of Fit Statistics Criteria* as follows:

Statistics	"Fit" Criteria
λ^2 - <i>Chie-Square</i>	p>0,05
<i>P-Value</i>	0,00
<i>RMSEA</i>	>0,05

(See Suhuri, 2009)

Based on the tryout results, it was shown that all load factor (λ) >0,3, *P-Value* was 0,00, and *RMSEA* >0,05. This result indicates that there is conformity between the speaking test instrument and the competence being evaluated.

Third Tryout (Implementation)

In reference to the above table, the validity of speaking test instrument administered at SMPN of West Bandung District is as follows:

Table 6 Validity of Speaking Test Instrument in Implementation Stage

Instrument	Load Factor Value (λ)	P-Value	RMSEA	df	Chie-Square	Validity
1	(λ)> 0,58	0,965	0,00	1.03	0,018	Valid
2	(λ)> 0,31	0,876	0,00	1.03	0,017	Valid
3	(λ)> 0,54	0,765	0,00	1.03	0,014	Valid
4	(λ)> 0,57	0,665	0,00	1.03	0,015	Valid
5	(λ)> 0,31	0,865	0,00	1.03	0,014	Valid
6	(λ)> 0,56	0,565	0,00	1.03	0,015	Valid
7	(λ)> 0,31	0,755	0,00	1.03	0,017	Valid
8	(λ)> 0,58	0,565	0,00	1.03	0,015	Valid
9	(λ)> 0,67	0,965	0,00	1.03	0,012	Valid
10	(λ)> 0,56	0,965	0,00	1.03	0,014	Valid
11	(λ)> 0,31	0,865	0,00	1.03	0,013	Valid
12	(λ)> 0,64	0,965	0,00	1.03	0,019	Valid
13	(λ)> 0,51	0,876	0,00	1.03	0,013	Valid
14	(λ)> 0,70	0,765	0,00	1.03	0,014	Valid
15	(λ)> 0,65	0,665	0,00	1.03	0,013	Valid
16	(λ)> 0,56	0,865	0,00	1.03	0,012	Valid
17	(λ)> 0,57	0,565	0,00	1.03	0,015	Valid
18	(λ)> 0,68	0,755	0,00	1.03	0,016	Valid
19	(λ)> 0,71	0,565	0,00	1.03	0,018	Valid
20	(λ)> 0,42	0,965	0,00	1.03	0,021	Valid
21	(λ)> 0,30	0,965	0,00	1.03	0,024	Valid
22	(λ)> 0,35	0,865	0,00	1.03	0,025	Valid
23	(λ)> 0,31	0,965	0,00	1.03	0,026	Valid

The above table shows that the validity of the 23 speaking test items administered at SMPN in West Bandung District was considered good with *Goodness of Fit Statistics Criteria* as follows:

Statistics	"Fit" Criteria
λ^2 - Chie-Square	p>0,05
P-Value	0,00
RMSEA	>0,05

(See Suhuri, 2009)

Based on the tryout results, it was shown that all load factor (λ) >0,3, P-Value was 0,00, and RMSEA >0,05. This result indicates that there is conformity between the speaking test instrument and the competence being evaluated.

DISCUSSION

As Table 3 indicates, the assessment model of Sundanese language speaking has 23 proficiency skills. The aspects analyzed include: (1) manual: clarity of scoring manual, (2) material: clarity of indicator, clarity of relationship between item type and indicator, relationship clarity between the instrument and the material being tested, and the clarity of evaluation material and testing instrument used; (3) language: use of standard language, definition of communicative language, use of easy-to-understand sentences and words; (4) writing procedures: font, font size, format or layout; and (5) grading guidelines of speaking test instrument.

Result of interview with experts shows that these 23 topics are comprehensive. They can also clearly describe different communicative purposes. As previous study has shown, when the participants apply these topics in their daily real life, they can get genuine response from their counterpart and get their communication objectives (Mally & Pierce, 1996).

Furthermore, two types of analysis were performed on the Indonesian language test instrument, namely qualitative analysis and quantitative analysis. Qualitative analysis is the analysis of speaking test according to expert judgment and user judgment, whereas quantitative analysis is performed by means of statistical method. Results of the analyses are described below.

The test manual for 23 test items of the Indonesian language speaking test was found to be the level of fair. This means that the existing manual needs to be improved. The directions need to be made in easy and quick to understand sentences, not too many, and readable. This is in line with the evaluation principle which considers that test or item directions are an important part of test development. Students will be able to do the test quickly if the directions are prepared and made well. No matter how good a test is, if the directions are unclear, the test will be of no good. Test directions will give directions to the students about what to do, how much time is given, the requirements that need to be fulfilled in doing the test, and what important thing the test takers need to know.

There are several important principles that need to be followed to make the test more effective. The first of these principles is certainty and clarity of what is being tested. A speaking test cannot be administered if the test objectives are not clearly and operationally defined. In order to evaluate students' learning progress, first we need to identify the learning objectives. Only then can we develop the evaluation instruments. Tests are the most common instrument of evaluation in education. A test should reflect the characteristics of the aspect to be measured. If we want to evaluate students' level of intelligence, then the components of intelligence and the learning achievement have to be clearly defined before a test is developed as an instrument of evaluation.

All the 23 items of the Indonesian language speaking test that were being developed were of fair quality. This means that the formulated indicators still needed to be modified. The indicator improvement that still needed to be made include the use of operational verbs, where some verbs were found to be less measurable and not an urgent indicator. For example in speaking test item 1, "*Mampu mendaftarkan pokok-pokok cerita pengalaman yang mengesankan* (Able to register the gist of an impressive experiential story)". This sentence was not a relevant indicator of the basic competence "*Mampu menceritakan pengalaman yang paling mengesankan dengan menggunakan pilihan kata dan kalimat yang efektif* (Able to recount/tell the most impressive experience using effective diction and sentences)". The most appropriate indicator that reflects the basic competence is "*Mampu menceritakan pengalaman yang paling mengesankan berdasarkan pokok-pokok rangkaian cerita dengan menggunakan pilihan kata yang tepat dan kalimat efektif.*" (Able to recount/tell the most impressive experience based on the main points of the story using effective diction and sentences.) The main or *the most important indicator that should be included in the speaking test is, therefore, "Mampu menceritakan pengalaman yang paling mengesankan berdasarkan pokok-pokok rangkaian cerita dengan menggunakan pilihan kata yang tepat dan kalimat efektif"* (Able to recount/tell the most impressive experience based on the main points of the story using effective diction and sentences). If this indicator can be achieved well by students, the indicator "Able to recount/tell the most impressive experience based on the main points of the story using effective diction and sentences" is measurable and, therefore, no longer need to be tested.

The material tested was explained on the basis of compatibility analysis with the basic

competence and indicator, material adequacy, being essential, real or realistic, and contextually appropriate in the use of Indonesian. The material being tested is reflected in the basic competence and indicator. An example from Speaking Test Instrument 1: The subject matter is “Story Telling” of the basic competence “Telling a story in a good order and appropriate voice, pronunciation, intonation, gesture, and and expression”. The test material that is relevant to the basic competence is therefore the aspects of order, voice, pronunciation, intonation, gesture and expression.

The scope of Indonesian speaking test material is reflected in the theoretical definition of speaking skill. Speaking skill is a complex linguistic skill which covers not only the question of pronunciation and intonation, but also the use of diction and other linguistic and non-linguistic elements. The aspects being evaluated in a speaking activity consists of linguistic and non-linguistic aspects. The linguistic aspects include: pronunciation, word stress, tone and rhythm, vocabulary or expression, sentence or sentence structure variations. The non-linguistic aspects include: fluency, material mastery, courage, friendliness, discipline, spirit and attitude.

Test material is linked to its context, therefore context plays an important role in communicative language skills and in the integrative approach to assessment vis a vis decontextual approach. Without (linguistic, discourse and sociocultural) context, language cannot have a full meaning. Oller (1979) says that a higher level of language is a contextualized perception of language. Achievement variability corresponds to the type of task involved, implied in the length of the test and type of text and adequacy of the format (see Douglas and Selinker, 1985; Skehan, 1987).

In estimating the ability to interact orally, we need to try to reflect on the interactive nature of a normal conversation and try to ensure that reciprocity is considered in the test task involved. The task has to be given under normal time and the unpredictable element of interaction should be identified in order that a genuine communication with other participants may take place.

A speaking test instrument should be considered from three points, namely type of evaluation, type of evaluation instrument, and format of evaluation instrument. Types of evaluation consist of process evaluation and product evaluation; types of evaluation instrument can be test and non-test. The format of evaluation instrument are oral format, written format, performance, project, portfolio, attitude (character) assessment, self-assessment, and observation. An example from speaking test instrument 4: the subject matter is “Delivering a story with props”, the basic competence “Story telling with props”, and indicator “Able to tell a story with props based on story points”. A non-test process evaluation in the format of performance and observation is used to evaluate the material, basic competence, and indicator. These types of evaluation, instrument and format are appropriate to evaluate the basic competence “Story Telling with Props” because speaking test instrument 4 has a logical connection with the indicator, basic competence and test material.

The advantage of speaking test format should take into account (a) the test size – length of the text, productivity, involvement; (b) grammatical complexity and scope of the required coherence devices; (c) functional scope – illocutionary variation levels involved; and (d) reference scope – the depth of knowledge relevant to the dictionary needed to handle the activity.

The language of a speaking test instrument has to have a standard grammar system in accordance with the standard grammatical rules. In addition to being grammatical, a speaking test also has to be written in sentences that are easy to understand or, in other words, readable, and communicative. Because a speaking test uses command and order sentences, it is not difficult to make the sentences. Here is an example from test item 21: *Berpidatolah dengan intonasi yang tepat serta artikulasi dan volume suara yang jelas!* (Make a speech with appropriate intonation and clear articulation and volume!) This test instrument sentence has covered the scope and aspects that students need to pay attention to.

A good speaking test instrument does not only have standardized language, it is also developed in a standardized format and writing system. The most dominant of this writing system is the spelling, such as the use of letters, the writing of words, and the use of punctuations and other written symbols. Because speaking test uses mostly command and order in its instruction,

the writing system and layout are relatively simple and easy, except when the test uses picture or sound stimulus.

Finally, it is interesting to note that the 23 assessment items of Sundanese oral proficiency developed have met the criteria of good quality assessment. Following Davies' (1977) criteria, the model has integrated listening and speaking skills into a communicative and authentic assessment. Besides, they can also help students get their communicative purposes. Having learned these oral proficiency test items at school, the participating students can apply this language competence within their authentic environment (Bachman, 1990; Morrow; 1977; Canale & Swain, 1980) such as schools and homes with parents and other family members. This authenticity helps students to learn the vernacular language more enthusiastically.

CONCLUSIONS

Based on the findings and discussion above, the study comes with some conclusions. First, statistical calculation shows that the twenty-three items of the Sundanese language speaking test are fit for use as a speaking skill evaluation instrument at Junior Secondary School (SMP) level. Results of tryout show that the assessment items conform to the theoretical construct that gives rise to the integrative test, i.e. a test that integrates linguistic competence with non-linguistic competence or competence and performance. Second, expert judgement also shows that these test instruments have clear guidelines for scoring, as well as clear relationship between the subject matter being evaluated and the evaluation instrument. Finally, tryout results also indicate that these speaking test instruments have been statistically shown to be consistent with the data found on ground.

REFERENCES

- Bachman, L.F. (1990). *Fundamental considerations in language testing*. Hong Kong: Oxford University Press.
- Borg, W.R., & Gall, M.D. (1983). *Educational research: An introduction*. (4th ed). New York & London: Longman.
- Brown, H.D. (2004). *Language assessment: Principle and classroom practices*. New York: Longman, Pearson Education, Inc.
- Canale, M. and M. Swain, 1980, 'Theoretical basis of communicative approaches to second language teaching and testing', *Applied Linguistics*, I: 1-47
- Carroll, B.J., 1980b, *Testing communicative performance: an interim study*, Oxford: Pergamon.
- Chaplen, E.F. 1970. "Oral examinations, in Centre for Information on Language Testing and Research, *Examination Modern Languages*. London: CILT and Committee on Research and Development in Modern Languages.
- Davies, A., 1977. 'The construction of language tests'. in Allen, J.P.B. and A. Davies (eds.). 1977: 38-104.
- Douglas, D. and L.Selinker, 1985, 'Principles for language tests within the "discourse domains" theory of interlanguage: research, test construction and interpretation', *Language Testing*, 2/2: 205-26.
- Fetterman, D.M. (1988). *Qualitative approaches to evaluation in education: The silent scientific revolution*. New York: Praeger Publishers.
- Haerudin, D., Suherman, A., & Nugraha, H. S. (2018). *The quality and values of character education in Sundanese language text book of the 2013 elementary school curriculum*. UPI 2nd International Conference on Language, Literature, Culture and Education (ICOLLITE 2018), 340-345.
- Hulin, C.L., Drasgrow, F., & Parsons, C. (1983). *Item response theory: Application*

- toppsychological measurement*. Homewood, Illinois: Dow Jones-Irwin.
- Johnson, D.W. & Johnson, R.T. (2002). *Meaningful assessment: A manageable and cooperative process*. Boston: Allyn and Bacon.
- Kelly, R., 1978. *On the construct validation of comprehension tests: an exercise in applied linguistics*, Ph.D. thesis, University of Queensland.
- Koswara, D., Dallyono, R., Suherman, A. & Hyangsewu, P. (2021). The analytical scoring assessment usage to examine Sundanese students' performance in writing descriptive texts. *Cakrawala Pendidikan*, 40(3), 573-583. DOI [10.21831/cp.v40i3.40948](https://doi.org/10.21831/cp.v40i3.40948)
- Kurniasih, N., Komariah, N. & Rodiah, S. (2014). Designing Sundanese cultural literacy programs based on national curriculum of 2013 for students of elementary schools in Cileunyi sub-district, Bandung Regency, Indonesia. *Proceedings 10th Biennial Convention of The Pacific and Asian Communication Association (PACA)*, Universitas Padjadjaran, 24-26 June 2014, 1735-1765
- Morse, J.M. (1994). *Critical issues in qualitative research methods*. London: Sage Publications.
- Mueller, R. O. (1996). *Basic principles of structural equation modeling: An introduction to LISREL and EQS*. New York: Springer.
- Morrow, K.E., 1977, *Techniques of evaluation for a national syllabus*, London: Royal Society of Arts.
- Muhtadi, D., Sukirwan, Warsito, & Prahmana, R. C. I. (2017). Sundanese ethnomathematics: Mathematical activities in estimating, measuring, and making patterns. *Journal on Mathematics Education*, 8(2), 185-198. DOI: 10.22342/jme.8.2.4055.185-198
- O'Malley, J.M. & Pierce, L.V. (1996). *Authentic assessment for English language learners: Practical approaches for teachers*. New York: Addison Wesley Longman, Inc.
- Oller, J.W., 1979, *Language tests at schools*, London: Longman.
- Porter, D., A. Hughes and C. Weir (eds.), 1988, *Validating the ELTS tests: a critical review*, Cambridge: British Council and UCLES.
- Skehan, P., 1987. 'Variability and language testing', in R. Ellis (ed.), 1987, *Second language acquisition in context*, Englewood Cliffs, NJ: Prentice Hall.
- Sudiyono, A. (2003). *Pengantar evaluasi pendidikan*. Jakarta: PT Raja Grafindo Persada.
- Sugiyono. (2004). *Metode penelitian bisnis*. Bandung: CV Alfabeta.
- Suhuri, 2009. *Model Evaluasi Pembelajaran Bahasa Inggris di SMA Palembang Tahun 2009*. (Disertasi). Yogyakarta: Pascasarjana UNY.
- Swain, M., 1985, 'Large-scale communicative language testing: a case study'. in Lee, Y.P. *et al.* (eds.), 1985: 35-146.
- Widdowson, H.G., 1978, *Teaching language as communication*, Oxford: Oxford University Press.
- Yusuf, Suhendra. 2006. *Pengembangan Model Ujian Nasional Berdasarkan Pendekatan Literasi: Kajian Struktur Soal dan Daya Serap Siswa SMP.MTs dan SMA/MA di Jawa Barat dalam Mata Pelajaran Bahasa Sundapada Ujian Akhir Nasional Tahun 2004*. (Disertasi). Bandung: Program Pascasarjana UPI.