Developing assessment instruments of debate practice in Indonesian Language learning

Septiana Farida*; Farida Agus Setiawati

Universitas Negeri Yogyakarta, Indonesia

*Corresponding Author. E-mail: sfarida2590@gmail.com

ARTICLE INFO

ABSTRACT

This study aims to develop an instrument for assessing debate practice in Indonesian Class X senior high school (*Sekolah Menengah Atas* or SMA/*Madrasah Aliyah* or MA) learning. The theoretical construct of the instrument was found after reviewing several theories, including speaking skills that apply to debate practice, especially those based on the Australian Debating Federation. The non-test instrument development procedure used is the Mardapi model, which includes non-cognitive. Ten material experts reviewed the draft instrument (two lecturers and eight Indonesian Language Teachers Class X in the SMA/MA in Yogyakarta Special Region) then it was calculated using the Aiken formula to prove the validity of the contents of the instrument. The draft instrument was also tested by two raters/evaluators to assess the debate practice. The results of this trial were used to calculate inter-rater reliability using Cohen Kappa. The assessment instrument was declared reliable from the calculation of the inter-rater reliability value of the Kappa formula, which was obtained at 0.678. The final item number of the instrument after the exploratory factor analysis is 33 items with adjustments to the composition of the dimensions of the statement items.

INTRODUCTION

Speaking skill is a second language skill acquired by humans before having reading and writing skills, which is practical oral communication carried out on every individual in the social environment (Simarmata & Sulastri, 2018). So far, speaking skills as part of communication have not been noticed, often ignored, and not taken seriously so that many students are unable and dare not speak (Isnaniar, 2013; Morelent, 2012). Speaking skills play a significant role in giving birth to a generation that is intelligent, critical, creative, and cultured (Isnaniar, 2013). In practice, speaking skills involve more complex aspects (Sari et al., 2016) and support other language skills (Simarmata & Sulastri, 2018).

The method used when speaking or in rhetoric is known as the art of speaking in dialogue or monologue. The art of speaking in the form of dialogue in question is a speaking activity that involves two or more people taking part in a conversation process (Midun, 2017, p. 14). The art form of speaking dialogue is debate, discussion, question and answer, negotiation, and conversation. The art form of monologue speech involves only one person speaking, namely in speeches, lectures, declamations, and remarks. Each speaking skills practice needs to be carefully studied and its components considered in every evaluation practice in the scope of Indonesian language learning in schools.

Speaking not only issues meaningless words, but it requires technique, clear thoughts, and content (Midun, 2017, p. 14). The technical components in question are breathing, voice building, reading, and storytelling techniques. Furthermore, clear and contained thoughts become part of the weight of the substance conveyed when speaking, namely whether they have high creative and fantasy power or knowledge and objective evidence (Midun, 2017, p. 2014). Therefore, learning speaking skills occupies an essential part (Isnaniar, 2013) and at the end of the learning requires a form of practical evaluation to observe all these components.

In principle, the implementation of the evaluation of language skills in schools takes place differently. Reading and writing skills are used in non-face-to-face communication, and the evaluation is done in writing through cognitive evaluation of the learning. The dominant mental evaluation is carried out and put forward by educators rather than affective or psychomotor evaluation. Cognitive evaluation is used as a benchmark for assessment and holds the principal place (Poerwanti et al., 2008, p. 23). This assessment is also evident from the national exam grid, which focuses on evaluating cognitive aspects at the elementary, junior high, and high school levels. The Indonesian national exam indicators only describe the evaluation of limited literary-non-literary reading and writing competencies and editing spelling (Badan Standar Nasional Pendidikan, 2018).

The form of evaluation of listening and speaking skills is carried out in practice when teaching is in progress (Isnaniar, 2013; Nurgiyantoro, 2001, p. 7). However, in reality, this skill evaluation is often forced in cognitive assessment through theoretical questions. If it is carried out in practice, it is realised without specific instrument guidelines. Speaking skills as a basic form of visual communication are often considered easy competencies, both to do and assess (Isnaniar, 2013; Morelent, 2012). The practice of evaluating speaking skills is not an easy thing to do (Sari et al., 2016, p. 2), and the form of evaluation is in the form of non-test instruments that can be in the form of observation sheets, questionnaires, or assessment rubrics and requires accuracy in the evaluation process.

Based on observations and unstructured interviews conducted with Indonesian teachers at MAN 3 Sleman, MAN 2 Kulonprogo, SMA N 6 Yogyakarta, and SMA N 9 Yogyakarta, the speaking skill assessment of students was carried out at a glance without special instruments. At a glance, here it is guided by the general aspects, namely intonation, expression, gesture, and mastery of the material, without elaborating the indicators in each of these aspects. These general aspects happen in every assessment of speaking practice, whether speech, sermon, negotiation, declamation, debate or drama. Each student certainly has different outstanding abilities between each component and deserves different values/weights of appreciation. In addition, based on the research mentioned by Brown (2015, p. 51) that fifteen of the sixteen students in the study commented that the use of debate in the classroom could improve collaborative skills or critical thinking skills during learning.

Debate is one of the arts of speaking in the form of dialogue learned in class X SMA/MA and does not yet have a structured assessment instrument. Therefore, assessment instruments are needed specifically made to support each form of speaking practice. Debate is a very complex speaking skill competency. In addition to involving a whole of personnel, the flow of the debate also requires adequate competence in speaking strategies. Speakers are not only necessary to be able to master the motions given so they can speak frankly and piously, but they must also can convince and be critical and at the same time break the opponent's opinion (Salim, 2015, p. 100). In addition, the debate also requires a reflective and neutral attitude and is critical in examining the arguments or evidence used. The debater must assess the problem with solid analysis, not just relying on the opponent's interpretation (O'Connor et al., 2018, pp. 90–91). Thus, the aspect of speaking skills in the practice of debate includes various components. This component becomes the basis for assessing the competence of each student when practising debate. In line with this, it is necessary to develop an instrument for assessing the competence of debating practice realized in a non-test instrument in the form of an observation sheet (Ghorbani et al., 2018).

The research by Viswesh et al. (2018) aims to evaluate students' ability to make evidence-based decisions and presentations through debate activities (methods). The results of this study indicate the readiness of team performance and students' skills in perceiving through debate. The process observed in this study is similar to the procedure followed in the development research carried out. The process in question is the various components and indicators that become the points of assessment in the debate, including the preparation process of the materials and methods of arguing used to obtain success in debating.

In addition, based on relevant theory and research results, it can be assumed that the instrument of debating practice ability consists of three factors: Matter, Method, and Manner. The problem is how to develop an instrument that can be used to assess the practice of debate in an appropriate (valid) and reliable (reliable) manner. Observation sheet to be filled out by the teacher when evaluating students' debating practices in class. Developing the instrument is carried out based on the theory of speaking that applies to the course of debate. A valuable instrument for assessing KD 4.12 Indonesian language learning in class X SMA/MA even semesters reads Based on the problems/issues, points of view and arguments of several parties and conclusions from verbal debates to show the essence of the debate.

The result of this study is the instrument used to assess the practice of debate that has good content validity and construct validity and has good interrater/Cohen Kapha reliability. The result of factor analysis (Exploratory Factor Analysis = EFA) shows that this debate instrument consists of three factors: Matter, Method, and Manner. The matter factor consists of 1 to 12 items, the method factor consists of 13 to 24 items, and the manner factor consists of 25-38 items. These three factors can explain the variance of debate practice by 100%.

METHOD

This study is a research on developing debate practice instruments using Djemari Mardapi's non-cognitive instrument development model (Mardapi, 2017). The assessment of debate practice in several debate contests that take place in the world after being studied from various sources refers to three dimensions: matter, method, manner (D'Cruz, 2003; Latif, n.d.; Quinn, 2005). The three dimensions are decided by the adjudicators or debate experts (jurors). Each size has a component description related to general speaking skills. This speaking skill material is used as an indicator of assessment. It is stated that the instrument statement on the components contained in each dimension of the debate.

The instruments that have been compiled were tested twice, namely limited trials and field trials. An expert judgment validation process preceded the trial against ten material experts in Indonesian language learning. After obtaining the validity value, the product was revised and tested limited (Murti, 2011, p. 20). A little trial was conducted on 24 students from four schools in Yogyakarta Special Region involving two assessors, namely the Indonesian language teacher from each school and researchers. The results of the limited trial obtained four values of inter-rater reliability using the Kappa formula calculation. Then, the average of the four reliability values was calculated, and the final inter-rater reliability value was obtained.

Furthermore, the same instrument was used in a field trial on 246 class X students from four schools in Yogyakarta Special Region. The results of the field trials obtained were analyzed by exploratory factors (EFA) to obtain construct validity. The product then underwent a second revision based on various outcomes of factor analysis and suggestions for modification, resulting in a final instrument. Moreover, the test subjects were students of class X, namely 24 students in the limited trial and 246 students in the field trial. The trial issue was determined by purposive sampling, namely the class X students who practised debating. Subject determination was assisted by Indonesian teachers from four different schools, namely MAN 3 Sleman, SMA N 6 Yogyakarta, SMA N 9 Yogyakarta, and MAN 2 Kulonprogo. The subjects of the field trial can be detailed as follows: 66 students of SMA N 6 Yogyakarta, 54 students of SMA N 9 Yogyakarta, 72 students of MAN 3 Sleman, and 54 students of MAN 2 Kulonprogo.

Septiana Farida & Farida Agus Setiawati

FINDINGS AND DISCUSSION

The product that resulted in this development research is an instrument sheet for assessing the practice of debate in Indonesian language learning for class X SMA/MA. The instrument sheet was developed to be used by Indonesian teachers in evaluating students when carrying out debating practices. The instrument sheet is in the form of a checklist observation sheet containing statements about the points that must be indicated to be observed when students argue.

In the initial development, instrument specifications were carried out, which included the preparation of statement items and poured into the instrument grid. The item statements totalled 44 statements. The grid was developed from three dimensions of debating practice assessment: matter, method, and manner. This dimension is determined by extracting from various literature on the evaluation of debate practice, including from international debate association institutions. Furthermore, each dimension is classified into components that give more specific indicators to be arranged into operational statement items.

The dimension of matter or material consists of three components: motions, arguments, and facts from statements. Measuring the method is also distinguished into three parts: delivering idea, submitting an objection, and providing the response. Dimensions of manner or attitude are classified into components of expression, appearance, and vocals. Each of these components is still classified into more specific indicators to be reduced to statement items. The naming of the elements in each dimension and the arrows for each component may still change concerning the results of the exploratory factor analysis based on field trial data.

The indicators derived from each component are reduced to item statements in more detail. Statements are coherently written on the instrument sheet according to the elements' order (Chai et al., 2019). The product instrument is a checklist observation sheet using a dichotomous score. These item statements are then observed in students when carrying out debate practices. The determination of the score on the observation sheet is carried out if the statement is found or observed, then the instrument sheet is given a check in the "Yes" column.

On the other hand, if the statement is not found or observed, the instrument sheet is marked with a tick in the "No" column. The checklist in the column also determines the score obtained by students because each statement rated "Yes" has a score of 1, and "No" has a score of 0. The total score of the observed items can be grouped into the categorisation of students' debating practice abilities. Then, from this categorisation, the score or predicate of students' level of proficiency in debating practice is known.

Content Validity

Validation of expert judgment is carried out before the product is used for testing. Expert validation was carried out on two Indonesian Language, and Literature Education Lecturers and eight Indonesian Language Teachers in class X. Expert validation aim to obtain content validity values calculated using the Aiken's V Index formula.

In the revision of the first stage, six items were dropped. These items do not pass the validity test because according to Aiken's Table, which refers to the number of raters, the value of the item validity coefficient must exceed 0.73. This value of 0.73 is the reference value for the 4-scale instrument at 10 raters with an error rate of 5%. The six items can be detailed as follows.

Table 1. Details of Dropped Items in the First Revision

Item Statement Number	Aiken's Validity Value	Indicator	Information
4	0.667	The substance of the motion	Safe to abort; other items represent indicators
13	0.700	Substance Facts	
17	0.333	Argument Statement	
21	0.600	Parry against Opponents	
29	0.700	Eye contact	
37	0.633	Costume	

Construct Validity

The developed instrument, which has been declared content valid and revised, is then used in a limited trial, besides that. The instrument is satisfactory and robust from Kappa's inter-rater reliability calculations. The instrument was tested on 246 students from four different schools. In this trial, each student debated their role as a pro and con group member and then directly assessed using the revised assessment instrument. The score data per item of the students in the form of a 1-0 dichotomous score were recapitulated and analysed using the SPSS program.

Results of Factor Analysis Items 1 to 12

Results of factor analysis items 1 to 12 represents the dimension of matter which consists of the components of motion, argument, and facts of the argument. Items are arranged coherently from each component. The motion component is divided into two indicators: the formulation of the motion with two statements and the substance of the motion with one idea. The argument component is divided into the essence of the argument with two item statements and each speaker's opinion with three-item ideas. Two indicators are components, including the identity of the facts with three statements and the substance of the points with one statement. The results of the KMO and Bartlett's Test and the total variance explained are shown in Figure 1 and Figure 2.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.628
Bartlett's Test of Sphericity	Approx. Chi-Square	2316.174
	df	45
	Sig.	.000

Figure 1. KMO for Items 1-12

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.551	35.514	35.514	3.551	35.514	35.514	3.388	33.879	33.879
2	1.987	19.873	55.388	1.987	19.873	55.388	1.943	19.435	53.314
3	1.638	16.377	71.765	1.638	16.377	71.765	1.759	17.586	70.899
4	1.010	10.104	81.869	1.010	10.104	81.869	1.097	10.970	81.869
5	.648	6.483	88.353						
6	.552	5.524	93.877						
7	.493	4.927	98.804						
8	.057	.575	99.379						
9	.051	.511	99.890						
10	.011	.110	100.000						

Extraction Method: Principal Component Analysis.

Figure 2. Total Variance Explained for Items 1-12
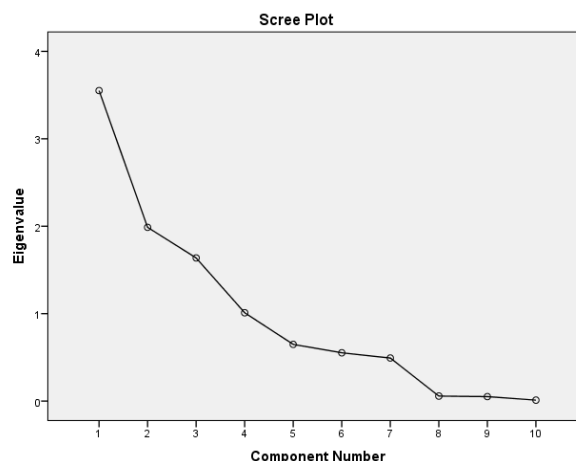
Figure 3. Scree Plot Display Dimension Matter

Septiana Farida & Farida Agus Setiawati

Table 2. The Naming of EFA Result Factors Dimension Matter

Component	Number of Statement Items	Percentage Variance	Factor Naming
1	9, 10, 11, 12	33.879	Fact of Argument
2	4, 5	19.435	Argument Statement
3	3, 6, 7	17.586	Contents of Speaker's Argument
4	2	10.970	Introduction to Arguments

From Figure 3, it is known that the analysis results obtained are that four factors have an Eigenvalue > 1. The following details are the values of the rotated component matrix. If the magnitude is > 0.5, it indicates a tendency to categorise the grain components. The four factors formed the group the statement items and led to naming the factors as presented in Table 2.

The analysis results show that the components developed in the instrument are primarily by the reality in the field, although there are improvements that need to be made. Point 1 can be aborted because apart from being represented by point 2, the "Motion Formulation" in debate practice has generally been formulated before the debate. Therefore, point 1 is not an item that must be observed in the implementation of debate practice because it is automatically present. Point 8 can be dropped because each speaker refers to the same concept of argument so that it cannot be separated, and each speaker strengthens the other speaker's argument. Therefore, arguments need not be restricted explicitly between speakers.

Results of Factor Analysis Items 13-24

Items 13-24 coherently represent the method's dimension, which consists of the components of how to convey arguments, how to submit rebuttals, and how to submit responses. The way the argument is delivered has two indicators: the argument statement with four-item statements and the speaker's opinion with two item statements. An indicator of providing defence is divided into two views of resistance items against opponent and two statements concerning the identification of reasons. Components of delivering responses are divided into objective responses with two item statements and response structure with two item statements.

Based on the factor analysis of the 12 item dimensions of the method, two statements with the smallest anti-image correlation value were produced, namely item 16 of 0.314 and item 13 of 0.473. The two items were aborted, and the other 10 items were re-analysed in the same way.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.841
Bartlett's Test of Sphericity	Approx. Chi-Square	1708.478
	df	45
	Sig.	.000

Figure 4. KMO for Items 13-24

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.995	49.948	49.948	4.995	49.948	49.948	4.977	49.769	49.769
2	1.241	12.406	62.355	1.241	12.406	62.355	1.241	12.412	62.181
3	1.080	10.801	73.156	1.080	10.801	73.156	1.097	10.974	73.156
4	.918	9.184	82.340						
5	.666	6.661	89.001						
6	.487	4.867	93.868						
7	.228	2.281	96.150						
8	.170	1.702	97.852						
9	.136	1.359	99.211						
10	.079	.789	100.000						

Extraction Method: Principal Component Analysis.

Figure 5. Total Variance Explained for Items 13-24

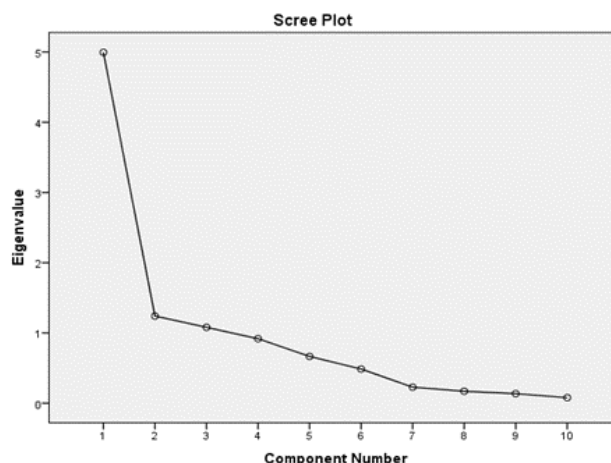Septiana Farida & Farida Agus Setiawati

Figure 6. Display of the Dimensional Scree Plot Method

Table 3. The Naming of EFA Result Factors Dimension Method

Component	Number of Statement Items	Percentage Variance	Factor Naming
1	18, 19, 20, 21, 22, 23, 24	49.769	How to Respond to Arguments
2	14, 15	12.412	How to Present an Argument
3	17	10.974	Arguing Rules

The KMO value generated after the second stage of analysis was 0.841. The results of the KMO and Bartlett's Test and the total variance explained are shown in Figure 4 and Figure 5. The scree plot in Figure 6 shows that there are three form factors which are indicated by the Eigenvalue > 1. The following is a breakdown of the rotated component matrix value; if the magnitude is > 0.5, it categorises the item component groupings. The three factors formed the group the statement items and led to naming the factors as shown in Table 3.

The concept of grouping indicators on each component of the developed method dimensions follows the factors formed in the results of the factor analysis carried out. The dominant factor of this dimension lies in how to convey responses, which in the development concept are divided into two indicators, namely responses and rebuttals. However, in the factor analysis results, these two indicators tend to group on one factor. Since rebuttal is also a form of response, factors are generally named ways of responding to arguments. The rest, items 14, 15, and 17, have occupied the same elements as the initial lattice development concept.

Results of Factor Analysis Items 25-38

Items 25-38 contain coherent statements from the manner dimension, composed of expression, appearance, and vocal components. Indicators of eye contact with one item statement, gestures with two item statements, and facial expressions with two item statements are descriptions of the expression components. The appearance component is detailed by one statement each of the indicators of standing and costume. The vowel component consists of voice and speed indicators with two item statements and pitch and pronunciation clarity which are also detailed in two item statements, respectively.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.719
Bartlett's Test of Sphericity	Approx. Chi-Square	945.433
	df	78
	Sig.	.000

Figure 7. KMO for Items 25-38

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.407	26.209	26.209	3.407	26.209	26.209	2.844	21.875	21.875
2	2.162	16.632	42.842	2.162	16.632	42.842	2.382	18.320	40.196
3	1.194	9.187	52.029	1.194	9.187	52.029	1.497	11.513	51.708
4	1.128	8.679	60.707	1.128	8.679	60.707	1.170	8.999	60.707
5	.894	6.880	67.587						
6	.819	6.301	73.888						
7	.701	5.392	79.279						
8	.613	4.714	83.994						
9	.565	4.344	88.338						
10	.543	4.175	92.513						
11	.494	3.797	96.310						
12	.368	2.832	99.142						
13	.112	.858	100.000						

Extraction Method: Principal Component Analysis.

Figure 8. Total Variance Explained for Items 25-38

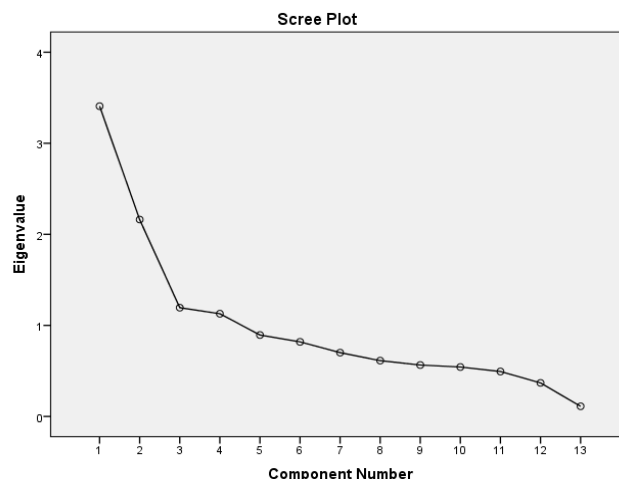

Figure 9. Scree Plot Display of Manner Dimensions

Table 4. The Naming of EFA Result Factors Manner Dimension

Component	Number of Statement Items	Percentage Variance	Factor Naming
1	32, 33, 34, 35, 36, 37	21.875	Vocal
2	25, 26, 27	18.320	Appearance
3	28, 29, 30	11.513	Expression
4	31	8.999	Costume

Based on the factor analysis of the 12 items of the Manner dimension, one statement was produced with the value of the rotated component matrix < 0.3, namely item 38 of 0.091. Thus, these items were dropped, and the other 11 things were re-analyzed with the same steps—items 25 to 37, analyzed by repeated factors, resulted in a KMO value of 0.719. The results of the KMO and Bartlett's Test and the total variance explained are shown in Figure 7 and Figure 8.

From Figure 9, it is known that four factors are formed, namely four points that exceed the Eigenvalue > 1. The grouping of items into four components can be seen in the classification and naming of factors in Table 4.

Items 32-37 were previously identified in the indicators specifically for the vocal component, but after factor analysis, it turned out that all of these items showed unity. Therefore, items 32-37 form a focused component as a vowel. Items 25 and 26-27 are different indicators, but they belong to the same component. However, after the factor analysis is carried out, it can be categorized into appearance components. Furthermore, item 30 in the initial development is a

Septiana Farida & Farida Agus Setiawati

different component of items 28-29. However, factor analysis tends to group the three items and can be categorized as expression components. Item 31 independently occupies the costume component factor because it is a statement that describes the costume.

Reliability

The instrument reliability value was obtained by calculating the Kappa Formula's inter-rater reliability average on 24 subjects from four different schools, namely from MAN 3 Sleman, SMA N 9 Yogyakarta, SMA N 6 Yogyakarta, and MAN 2 Kulonprogo. The assessment results between the two assessors from the four schools were categorized first and then analyzed for their Kappa scores with the help of the SPSS Program. The categorization is determined into five groups based on the categorization guide Azwar (2012, p. 140), namely by first calculating the minimum value, maximum value, range, mean, and standard deviation.

The instrument is in the form of an observation checklist with 38 statements whether or not there is so that the minimum value is 0 and the maximum value is 38. The range or range between the maximum and minimum values is 38. The mean is the maximum and minimum value divided by two, which is 19. The standard deviation is obtained from the range separated by six, which is 6.3. Based on these benchmarks, the categorization as shown in Table 5 can be obtained.

Table 5. Categorization of Students' Debate Practice Ability

Predicate	Value Range	Nominal Score
Ver Low	$X \leq 9.55$	1
Low	$9.55 < X \leq 15.85$	2
Medium	$15.85 < X \leq 22.15$	3
High	$22.15 < X \leq 28.45$	4
Very High	$X > 28.45$	5

There are 24 subjects divided into four schools obtained a nominal score categorisation ranging between 4 and 5. Kappa reliability values obtained from data from MAN 3 Sleman, SMA N 6 Yogyakarta, and SMA N 9 Yogyakarta were 0.571. In contrast, the reliability value Kappa data obtained from MAN 2 Kulonprogo is perfect, which is 1. The Kappa reliability value of the assessment instrument developed is calculated from the average of these four values. The average weight of Kappa reliability obtained is 0.678. Based on the categorisation of the Kappa reliability value by Fleiss and Cohen (1973), it can be seen that the debate practice assessment instrument developed is in the sufficient predicate because it is in the range of 0.61 to 0.75. It was corroborated by Garson (2016, p. 65), who also stated that the Kappa inter-rater reliability value was between 0.6 to 0.79 is included in the substantial (sturdy) category. Therefore, it can be said that the instrument developed is reliable.

CONCLUSION

The research and development that has been carried out have produced an instrument for assessing the practice of debate in Indonesian language learning for students of class X SMA/MA, tested for validity and reliability. Before further research, the theoretical construct of the instrument was found after examining several theories about speaking skills in debate practice. After that, the procedure for developing the non-test instrument used in the study used the Mardapi model, which included non-cognitive. The draft instrument has been assessed by 10 experts, two of whom are dialectical speaking skills lecturers at the Indonesian Language and Literature Education Study Program and eight Indonesian Class X teachers in the SMA/MA in Yogyakarta Special Region. Both of them conducted an assessment using the Aiken formula to prove the validity of the instrument content. The initial instrument developed consisted of 44 items. However, after operating field trials and conducting factor analysis, 33 final instruments were produced.

The research results are as follows. First, the product of the debate practice assessment instrument compiled has been tested and has an adequate content validity value. The content validity test with the Aiken's V Index formula produces a validity value of 0.73 and is classified as valid. The first product revision was carried out after this calculation, which was to abort six statements so that the instrument totalled 38 items from the previous 44 items. Second, the product reliability value is included in the reliable category, which is 0.678. Reliability is generated from the calculation of inter-rater reliability using the Kappa coefficient, which is based on the average of the four inter-rater reliability values. The four inter-rater reliability scores were obtained from a limited trial of class X students from four different schools. Thus, the number of statement items for the debate practice assessment instrument developed after a construct validity test with exploratory factor analysis (EFA) is 33 items originating from three assessment dimensions. The three dimensions include the Matter dimension with four components/10 item statements, the Method dimension with three elements/10 item statements, and the Manner dimension with four components/13 statement items.

ACKNOWLEDGMENT

REFERENCES

Azwar, S. (2012). *Reliabilitas dan validitas* (4th ed.). Pustaka Pelajar.

Badan Standar Nasional Pendidikan. (2018). *Kisi-kisi USBN dan UN*. Badan Standar Nasional Pendidikan. https://bsnp-indonesia.org/2018/11/bsnp-rilis-kisi-kisi-usbn-dan-un-2019/

Brown, Z. W. (2015). The use of in-class debates as a teaching strategy in increasing students' critical thinking and collaborative learning skills in higher education. *Educationalfutures [Online]*, *7*(1). https://educationstudies.org.uk/?p=3685

Chai, C. S., Hwee Ling Koh, J., & Teo, Y. H. (2019). Enhancing and modeling teachers' design beliefs and efficacy of technological pedagogical content knowledge for 21st century quality learning. *Journal of Educational Computing Research*, *57*(2), 360–384. https://doi.org/10.1177/0735633117752453

D'Cruz, R. (2003). *The Australia-Asia debating guide* (2nd ed.). The Australian Debating Federation. https://www.dav.com.au/resources/aadg.php

Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, *33*(3), 613–619. https://doi.org/10.1177/001316447303300309

Garson, G. D. (2016). *Partial least squares: Regression & structural equation models*. Statistical Publishing Associates.

Ghorbani, S., Mirshah Jafari, S. E., & Sharifian, F. (2018). Learning to be: Teachers' competences and practical solutions: A step towards sustainable development. *Journal of Teacher Education for Sustainability*, *20*(1), 20–45. https://doi.org/10.2478/jtes-2018-0002

Isnaniar. (2013). *Peningkatan kemampuan berbicara siswa kelas XI SMA Negeri 4 Kota Bengkulu tahun ajaran 2012-2013 dengan pendekatan komunikatif* [Universitas Bengkulu]. http://repository.unib.ac.id/id/eprint/8515

Latif, M. A. (n.d.). *A comprehensive guide to debate adjudication*. International Islamic University Malaysia. http://phased-uph.weebly.com/uploads/3/2/1/6/32162939/comprehensive_adjudication_guide.pdf

Mardapi, D. (2017). *Pengukuran, penilaian, dan evaluasi pendidikan* (2nd ed.). Parama Publishing.

Midun, H. (2017). Membangun budaya mutu dan unggul di sekolah. *Jurnal Pendidikan Dan Kebudayaan Missio*, *9*(1), 50–59. http://unikastpaulus.ac.id/jurnal/index.php/jpkm/article/view/117

Morelent, Y. (2012). *Peningkatan kemampuan berbicara siswa melalui kegiatan bercerita berbasis karakter di Sekolah Menengah Atas: Studi kuasi eksperimen pada siswa kelas X SMA Banuhampu Kabupaten Agam* [Sekolah Pascasarjana Universitas Pendidikan Indonesia]. http://repository.upi.edu/7716/

Murti, B. (2011). Validitas dan reliabilitas pengukuran. In *Matrikulasi Program Studi Doktoral, Fakultas Kedokteran, UNS*, 1-19. https://dokumen.tips/documents/validitas-reliabilitas-pengukuran-prof-bhisma-murti-55cd8744673e9.html?page=19

Nurgiyantoro, B. (2001). *Penilaian dalam pengajaran bahasa dan sastra*. BPFE-UGM.

O'Connor, A., Carpenter, B., & Coughlan, B. (2018). An exploration of key issues in the debate between classic and constructivist grounded theory. *Grounded Theory Review*, *17*(1). http://groundedtheoryreview.com/2018/12/27/an-exploration-of-key-issues-in-the-debate-between-classic-and-constructivist-grounded-theory/

Poerwanti, E., Widodo, E., Masduki, Pantiwati, Y., Poerwanti, E., Widodo, E., Masduki, Pantiwati, Y., & Departemen Pendidikan Nasional. (2008). *Asesmen pembelajaran SD*. Direktorat Jenderal Pendidikan Tinggi Departemen Pendidikan Nasional.

Quinn, S. (2005). *Debating*. Simon Quinn. https://debate.uvm.edu/dcpdf/quinn_DEBATING.pdf

Salim, A. (2015). Debate as a learning-teaching method: A survey of literature. *TARBIYA: Journal of Education in Muslim Society*, *2*(1), 97–104. https://doi.org/10.15408/tjems.v2i1.1665

Sari, K. D. I., Wendra, I. W., & Wisudariani, N. M. R. (2016). Pelaksanaan evaluasi pembelajaran keterampilan berbicara (bercerita) dengan materi cerpen pada siswa kelas IX D SMP Negeri 3 Singaraja. *Jurnal Pendidikan Bahasa Dan Sastra Indonesia Undiksha*, *5*(3). https://ejournal.undiksha.ac.id/index.php/JJPBS/article/view/8688

Simarmata, M. Y., & Sulastri, S. (2018). Pengaruh keterampilan berbicara menggunakan metode debat dalam mata kuliah Berbicara Dialektik pada mahasiswa IKIP PGRI Pontianak. *Jurnal Pendidikan Bahasa*, *7*(1), 49-62. https://journal.ikippgriptk.ac.id/index.php/bahasa/article/view/826

Viswesh, V., Yang, H., & Gupta, V. (2018). Evaluation of a modified debate exercise adapted to the pedagogy of team-based learning. *American Journal of Pharmaceutical Education*, *82*(4), 345–353. https://doi.org/10.5688/ajpe6278