

# Innovative Assessment Instrument to Produce Report Text for Madrasah Students

Shifa Fauziah<sup>1,\*</sup> Subyantoro Subyantoro<sup>2</sup>, Haryadi Haryadi<sup>3</sup>, Hari Prastyo<sup>4</sup>,  
Desi Wulandari<sup>5</sup>, Erna Sundari<sup>6</sup>

<sup>1</sup> MAN 1 Jepara Ministry Religion of Central Java Jepara, Indonesia

<sup>2</sup> Faculty of Language and Art, Universitas Negeri Semarang, Semarang, Indonesia

<sup>3</sup> Faculty of Language and Art Universitas Negeri Semarang, Semarang, Indonesia

<sup>4</sup> Islamic Education Department Institut Agama Islam Uluwiyah Mojokerto, Indonesia

<sup>5</sup> Postgraduate Science of Education Universitas Negeri Surabaya, Surabaya, Indonesia

<sup>6</sup> MTsN 11 Tasikmalaya, Ministry Religion of west Java, Tasikmalaya, Indonesia

\*Corresponding author. Email: shvfauziah339@gmail.com

## ABSTRACT

The demands of language learning are not only about the vocabulary and style but also the aspects of communication and cultural ethics that make assessment instruments need to be innovated. Innovation in this research is a development assessment instrument to produce report text with Valette's Taxonomy and CLIL approach. The purposes of this research are to describe the characteristics, innovation, validity, and reliability of an assessment instrument to produce report text for *Madrasah Aliyah* (Islamic Senior High School) students. The research method used in this study was research and development which was adapted from Borg and Gall (1983) with seven-step research; a preliminary survey, initial data collection, design of products, validation of products, revision and improvement of design, limited testing, revision and improvement of design. The results revealed that the characteristics assessment instrument to produce report text for *Madrasah Aliyah* students covered four main objectives; language, culture, communication, and literature. The development assessment instrument included an improvement on the cover, syllabus, lesson plans, instruction for questioning, and question items. The validity and reliability of the assessment instrument showed that four multiple-choice questions, two short answer questions, and an essay were invalid so it needed to be repaired.

**Keywords:** *Assessment instrument, Content language integrated learning, Madrasah aliyah, Report text, Valette's taxonomy.*

## 1. INTRODUCTION

The paradigm that "*language as a carrier of knowledge*" has made fundamental changes to the objectives of Indonesian language learning. Indonesian language learning is not only about knowledge or communication skills but also about a text that serves as a source of self-actualization in a particular socio-cultural context. It is important to face society 5.0 that the ability of learners to actualize themselves in a society with different cultural backgrounds is needed.

Some facts show the achievement of language learning goals is still far from expectations. One of the facts of the PISA (Program for International Student Assessment), the survey conducted by OECD in 2018 shows that reading literacy Indonesia is in 74<sup>th</sup> place from 79 countries or in the fifth-lowest position [1]. A low

level of literacy causes low problem-solving skills to the learners. This means that language learning has not been able to build students' way of thinking. In terms of communicating, language learning functions as a means of forming students' cognition.

Many factors need to be addressed to improve the achievement of language learning goals. One of which is the assessment or evaluation system. The achievement of language learning objectives requires an assessment instrument that is not only able to measure learners' understanding of grammar, skills of student's communication, and critical thinking of learners in response to information, but also students' sensitivity to social phenomena and the cultural context.

Report text is one of the texts listed in the text-based curriculum for high school students. The learners' ability

to produce report text can be used to collect and communicate the learning outcomes they have acquired [2]. Measuring the ability of learners in producing report text requires assessment instrument that not only tests the learners' knowledge about the structure and language elements but also the main idea of a paragraph. The learners also should be able to understand and use the text along with the cultural context [3].

In text-based curriculum, mentioned as *Wahana Pengetahuan* [4], the inclusion of context situation and culture is clearly appropriate on report text. However, it is unfortunate if the indicators assessment instruments do not involve the context situation and culture. The indicators of context situation and culture should be involved in the assessment of learners' exercises, the assessment of learners' activities, formative and summative assessment, and assessment of learning progress through the portfolio.

Based on the results of the analysis assessment instrument used in the *Madrasah Aliyah Negeri 1 Jepara* (Islamic Public Senior High School), it was found that the test is still in the form of questions requiring theoretical knowledge of text structure and grammar. The problems posed to the learners are also not following the context situation and culture. Most of areas to test the language proficiency of learners are the skills of reading, writing, and speaking. The rest is just a few questions to test other competencies, such as linguistic, socio-cultural, and strategy competencies. The student activity assessment rubric is also made as concise as possible without even including clear scoring criteria.

Based on interviews with some teachers at *Madrasah Aliyah Negeri 1 Jepara*, it was revealed that the teachers need clear and specific assessment guidelines following the material to be taught. The teachers found that it is difficult to formulate different assessment indicators in each subject that is going to be taught to the learners. There are many aspects of the assessment that must be completed in the text-based curriculum. This is one of the obstacles to teacher assessment. Although the cultural aspect and the situational context are integrated with the material, the assessment instrument of the cultural aspect should be given a special space by formulating the indicators' form independently with due regard to the practicality and completeness of the assessment aspect.

These facts show that the improvements of assessment instruments in language learning are necessary. The position of language learning, especially the Indonesian language, as a carrier of knowledge for other subjects, makes language learning has different characteristics than other subjects. This means the taxonomy of language learning should be distinguished from other taxonomic subjects.

Valette stated that in addition to exploring the three aspects of attitude, knowledge, and skills, taxonomy also

has simple content and complete content that has been adapted to the needs of language learning [5]. The content of the material is spelled out in eight areas and grouped into four main objectives. The first objective is the purpose of the language including spoken language, written language, and body language or kinesics. The second objective is a cultural goal that includes the way of life in a cultured society, knowledge of civilization, and art. The third goal is literacy, and the fourth goal is communication [6].

The presence of internal and external domains at every level is another advantage of Rebecca M. Valette's learning taxonomy. Internal domains are used to measure the attainment of receptive skills of learners (reading and listening) and the external domains are used to measure the productive skills of learners (writing and speaking).

In addition to the learning taxonomy, learning approaches must be considered to achieve the language learning goals. Nowadays, content-based language or content-based learning is the learning approach needed. The appropriate approach is the approach of Content and Language Integrated Learning (CLIL). This approach integrates language learning with the teaching of other academic subjects [7]. In CLIL, learners are allowed to learn and think about how to use the language, not only thinking about language as the main focus of learning. Coyle [8] and Mehisto [9] mentioned four key principles of CLIL: content, communication, cognition, and culture. Therefore, the CLIL approach is particularly suitable when combined with Rebecca M. Valette's taxonomy as both have similar content.

The integration of CLIL approach and learning taxonomy can achieve language learning objectives because they have cultural and communicative content which are needed in report text. The four basic components of the CLIL approach will be able to support the four aspects of Rebecca M. Valette's taxonomy. Both have harmonious and complementary content and objectives.

The purposes of this research are to describe and develop the characteristics of an assessment instrument to produce a report text for *Madrasah Aliyah* (Islamic Senior High School) students, and to test the validity and reliability of assessment instrument to produce a report text for *Madrasah Aliyah* students.

## 2. METHODS

This research used the research and development approach, adapted from Borg and Gall [10] with seven steps Research; preliminary survey, Initial data collection, product design, product validation, revision and design improvement, limited trials, revisions and improvements design. The stages of this research can be seen in figure 1.

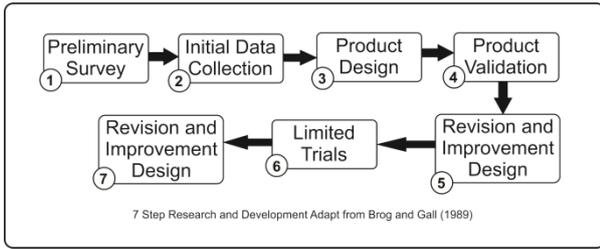


Figure 1. The Stages of Research.

The data in this research included the tendency score of the assessment instrument according to teachers' perceptions, the measurement of assessment instrument by experts, and the effectiveness assessment instrument to produce report text for Madrasah Aliyah.

The score of the evaluation report was obtained from first year Indonesian teachers from the Madrasah Aliyah Negeri 1 Jepara, Madrasah Aliyah Negeri 3 Klaten, and Madrasah Aliyah Negeri 2 Banyumas (all are Islamic Public Senior High School). The experts' assessment scores were obtained from an expert in learning evaluation, a linguist, and two language teachers in Indonesia. The effectiveness score of the assessment instrument was obtained from 32 participants in first-year students at Madrasah Aliyah Negeri 1 Jepara.

This research used a structured questionnaire and interview. The structured questionnaire was used to

Table 1. Characteristic of assessment instrument to produce report text for madrasah aliyah

Aspects		Characteristic of Assessment Instrument
<i>Language Objectives</i>		
1	Cognitive assessment of spoken language to build a positive attitude in responding phenomenon.	Question with 3 paragraphs of text as context, accompanied by multiple-choice and the answer key containing criteria. The duration is about 40 minutes for 10 questions and the level of difficulty is 30% easy-40% middle-30% difficult.
2	Cognitive assessment of rules and pattern written language to appreciate and practice the thought of religion.	Question with 3 paragraphs of text as context, accompanied by multiple-choice and the answer key containing criteria. The duration is about 30 minutes for 10 questions and the level of difficulty is 30% easy-40% middle-30% difficult.
3	Skill assessment of identifying rules and patterns is written language to build a positive attitude in responding phenomenon.	Question with 3 paragraphs of text as context, accompanied by project, 4 indicators, and 4 score range of assessment. The duration is about 2x40 minutes.
4	Skill assessment of producing speech and pattern to convey the implicit intent of the spoken language.	Question with 4 paragraphs of text as context, accompanied by performance test, 5 indicators, and 3 score range of assessment. The duration is about 2x40 minutes.
<i>Cultural Objectives</i>		
5	Cognitive assessment of understanding the explicit meaning of socio-cultural patterns to appreciate and practice the thought of religion.	Question with 3 paragraphs of text as context, accompanied by a short answer and the answer key containing criteria. The duration is about 30 minutes for 10 questions and the level of difficulty is 25% easy-50% middle-25% difficult.
6	Skill assessment of produce socio-cultural elements and patterns to have a	Question with 4 paragraphs of text as context, accompanied by performance test, 5 indicators, and 6 score range of assessment. The duration is about 1x40 minutes.

capture the tendency of the assessment instrument. The interview was addressed to the teacher. Validation was directed to the expert. The validity and reliability were shown from the instrument for scoring the data. The effectiveness of the assessment instrument was shown from trial, interview guides, sheets observation, and documentation.

### 3. RESULTS AND DISCUSSIONS

Results of this research shows the characteristics, development, validity and reliability of an assessment instrument to produce report text for Madrasah Aliyah students.

#### 3.1 Characteristic of Assessment Instrument to Produces Report Text for Madrasah Aliyah Students

Based on teachers' perceptions, the assessment instruments to produce report text included the four objectives: language objectives, cultural objectives, literary objectives, and communication objectives. The four objectives were developed into the characteristic of an assessment instrument to produce report text for Madrasah Aliyah students. Table 1 shows the characteristic of assessment instruments to produce report text for Madrasah Aliyah students.

Aspects		Characteristic of Assessment Instrument
	continuous desire in improving competence and understanding.	
<i>Literary Objectives</i>		
7	Cognitive assessment of literature to motivate in the responding phenomenon.	Question with 3 paragraphs of text as context, accompanied by multiple-choice and the answer key containing criteria. The duration is about 30 minutes for 10 questions and the level of difficulty is 25% easy-50% middle-25% difficult.
8	Skill assessment of evaluating literary phenomenon to build a positive attitude.	Question with 4 paragraphs of text as context, accompanied by performance test, 5 indicators, and 6 score range of assessment. The duration is about 1x40 minutes.
<i>Communication Objectives</i>		
9	Cognitive assessment of analyzing utterances or implicit patterns of communication to appreciate and practice the thought of religion.	Question with 3 paragraphs of text as context, accompanied by multiple-choice and the answer key containing criteria. The duration is about 40 minutes for 10 questions and the level of difficulty is 25% easy-50% middle-25% difficult.
10	Skill assessment of producing utterances or patterns to convey the explicit meaning of communication.	Question with 4 paragraphs of text as context, accompanied by performance test, 5 indicators, and 6 score range of assessment. The duration is about 1x40 minutes.

In cognitive assessment, there were several aspects discussed: a form of assessment, number of paragraphs, assessment constraints, duration of the assessment, and difficulty level of assessment. While on the skill assessment aspect, there were form of assessment, the number of paragraphs, the number of indicators of assessment, the duration of assessment time, and the range assessment.

The first aspect discussed in the characteristic of language objectives was the form of the test. Teachers chose multiple-choice questions with reasons of easiness in scoring, and the time was required is shorter. Teachers could also find out the test results quickly, and the information obtained can be more varied. Results of this research were strengthened by Wirayasa [11] which the advantages of multiple-choice questions were the items that can be used to measure all levels of knowledge, the characteristics of tests that required a relatively short working time, and scoring that could be done objectively [12] [13].

In an assessment of skill competence, there were three choices of assessment: performance, project, and portfolio. Teacher choice fell on performance and project. The assessment of the portfolio was complicated because it had a lot of criteria, so that it felt incriminating to teachers. Portfolio assessment criteria included documenting assessment results in the prescribed format. The teachers' difficulties are documenting each learning outcome of learners, taking time and special perseverance to realize a complete and neat learning result document. Besides, not all learners had the diligence and thoroughness to collect tasks that had been

done. The result of this research was contradictory to Pourdana's research [14]. Pourdana's research explored that a genre-based portfolio could impact descriptive and narrative writing performance. Students could engage, perceive, and act upon teacher feedback. This result also contradictory to Kovacek & Bode asserted that a portfolio was aimed to monitor the students' improvement in a certain period [15].

The second aspect discussed was the number of paragraphs. The number of paragraphs selected by the teacher was good on the cognitive or skill assessment that was between 3 to 4 paragraphs of each text. The selection of paragraph numbers by teachers was based on the reason that the text of the observational report had three parts: general definition, section description, and benefit description [16].

The third aspect discussed was the limits for cognitive assessment and the number of indicators for skills assessment. On multiple-choice questions, the teacher chose the answer key, on short answers and essays teachers selected the answer criteria. This is in line with Rauf [17] revealing that multiple-choice questions were the most common assessment tool used to assess cognitive knowledge. As for the number of indicators used as a guideline on skills assessment, teachers chose to use 3-4 indicators so as not to be so long and difficult.

The fourth aspect discussed was the duration of the assessment. Teachers chose to spend 40 minutes working on ten multiple-choice questions, short answers, and descriptions. For multiple-choice questions, solving ten questions within 40 minutes was possible. However, for questions and short answers, 40 minutes was less. This is

in line with Mistra [18] [19] mention that the amount of time spent was believed to aid productivity and help the student to complete tasks on schedule.

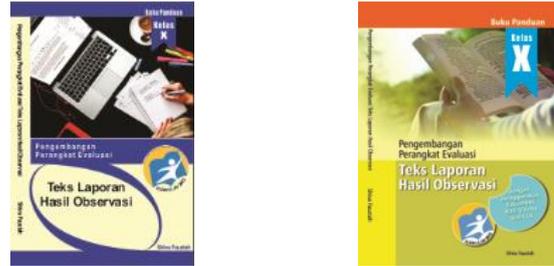
The fifth aspect was the level of difficulty of the cognitive assessment and the number of scales of assessment on skills assessment. As for the difficulty level, teachers chose the difficulty level of 30% about easy-40% about medium-30% difficult questions. This is following the theory of the level of difficulty which has an opportunity to correctly answer a question at a certain level of ability. The ratio of moderate-to-moderate problem questions and the difficult questions used in theory was 1: 2: 1 [20]. It is important to examine the level of difficulty to figure out the quality of the test [21]. The range of skill assessment aspect was adjusted with teacher requirement that is 3-4 range [22].

### **3.2 Innovative Assessment Instrument to Produce Report Text for Madrasah Aliyah Students**

The assessment instrument of innovates was a book containing an assessment instrument for evaluating the competency of producing report text for *madrasah aliyah*. The basic competence (*Kompetensi Dasar* or *K.D*) of producing report text is stated in K.D 3.1, 3.2, 4.1, and 4.2 for the first-year students of *madrasah aliyah*. This basic competence was combined with Valette's taxonomy and Context Language Integrated Learning (CLIL) approach.

There were two steps in innovating an assessment instrument's book. The first was the prototype product and the second was the final product. A prototype product is a product developed based on the characteristic of the teacher's perception. The prototype product consisted of five chapters. The first chapter was about assessment instruments, the second chapter was about report text assessment instrument, the third chapter was about the principle of the assessment instrument, the fourth chapter was about assessment instrument of cognitive, and the fifth chapter was about assessment instrument of skill. Attitude assessment was integrated with the fourth and the fifth chapter including cognitive and skill assessment.

The prototype product then was evaluated by the teachers and experts to produce the final product. Feedback suggestions given by the teachers and expert lecturers were used for product improvements. Improvements were made on some cover components, such as on the front cover and back cover. Repairing the front cover was replacing and adding images that match the theme of the assessment instrument, as well as rearranging layout design. On the back cover, the repair was layout some images. The following is a cover of the prototype product and final product.



**Figure 2.** Cover of prototype product and the final product.

In the prototype product, the learning syllabus had not been presented. However, as the recommendation from the experts and teachers, the syllabus was presented as the final product to determine the overall picture of the material on the assessment instrument. This is in line with Torres [23] revealed that syllabus was important for guidance to organize the lesson and help students to achieve quality in teaching and learning.

There were two improvements in the implementation of the lesson plan; the addition of tables in the assessment and recap assessment results. The addition of tables in the assessment section was intended to facilitate the teacher in understanding the techniques and forms of assessment to be used. Recapitulation of assessment results also aimed to facilitate teachers in the calculation of assessment of learning outcomes [24].

Following the advice from the teachers and experts, in the preparation of workmanship, instructions should use sentence suggestions rather than command sentences. Therefore, the command phrase in the manual instruction was replaced with a suggested sentence for example, "Learners can choose the most appropriate answer by cross-marking (X) in letters a, b, c, or d". It aims to increase the value of politeness on assessment instruments [25]. In addition to improvements to the manual problem, this suggestion was also applied to guidance on project tasks and performance on skills assessment.

There were improvements to linguistic aspects of multiple-choice items, short answers, and essays. The improvements were especially on items that contain the word "you/us". We omitted the words "you and us" on the item to avoid the subjectivity of the answer.

### **3.3 Validity and Reliability of Assessment Instrument to Produces Report Text for Madrasah Aliyah**

The validity and reliability were carried out in two stages. It consists of 20 multiple choice items, 10 short answers, and 5 essays. The 20 multiple-choice questions were tested on 32 students of first-year students. It was obtained four invalid questions that were on the question of numbers 7,10,12, and 13. Causes of invalidity is due to and the value of r arithmetic <value r table (N = 32; 5

$\%) = 0.349$ . A valid item was 16 questions. The reliability question was tested for 0.658. For the level of difficulties, there were 5 questions which had an easy level, 8 questions medium, and 7 questions difficult. The strength of difference with good criteria is as much as 5 questions, enough criteria as much as 6 questions, and bad criteria as much as 9 questions. Based on these results, this requires improvements in the evaluation tool so that the assessment can be more valid and reliable. Improvements are made by presenting new problems back to learners.

The result of a multiple-choice test of the second stage shows that the 20 multiple-choice question is entirely valid. The whole problems are valid because the value of  $r > r$  table value ( $N = 32; 5\%$ ) = 0.349. Problem reliability increased to 0.755. The difficulty level 10 questions on easy criteria, 6 questions on medium criteria, and 4 questions on difficult criteria. The strength of difference, 2 questions with excellent category, 3 questions with good category, 6 questions with enough category, and the remaining 9 problems with bad category [26].

The 10 short answer questions are tested on the 32 students of first-year students of Madrasah *Aliyah*. It is obtained 2 invalid questions that are on the numbers 2 and 4. The cause of the invalidity of the problem because and the value of  $r$  arithmetic  $< \text{value } r \text{ table } (N = 32; 5\%) = 0.349$ . A valid item of 8 questions. The reliability question is tested for 0.443. The difficulty level, 1 question with medium category, and 9 others have an easy category. The strength of difference with excellent criteria is 3 questions, good criteria is 1 problem, enough criteria is 1 question, and bad criteria is 5 questions. Based on these results, the assessment needs improvements in the evaluation tool so that more valid

and reliable. The improvements are made by presenting new problems back to learners.

The result of the second test of the short answer shows that the acquisition of 10 short answer questions is entirely valid. The whole problems are valid because the value of  $r$ -arithmetic  $> r$ -table value ( $N = 32; 5\%$ ) = 0.349. The reliability increased to 0.528. The difficulty level about the 9 questions remains on easy criteria and 1 question is medium criteria. The strength of difference, 3 questions with good category, 1 question with enough category, and 6 questions with bad category.

From 5 essay questions tested to 32 students of the first-year students of Madrasah *Aliyah*, 1 question is invalid because of the matter and value of  $r$  count  $< \text{value } r \text{ table } (N = 32; 5\%) = 0,349$ . 4 questions are valid items. The reliability question is tested for 0.248. The difficulty level, all with medium category. The strength of difference with excellent criteria is as much as 1 question, and enough criteria is 4 questions. Based on these results, the assessment needs improvements in the evaluation tool so that more valid and reliable. The improvements are made by presenting new problems back to learners.

The result of the second essay test shows the acquisition of five essays is entirely valid. The whole problems are valid because the value of  $r$  arithmetic  $> r$  table value ( $N = 32; 5\%$ ) = 0.349. Problem reliability increased to 0.325. The difficulty level of about 5 questions remains on the medium criteria. The strength of difference, 1 question with good category, and 4 questions with enough category. The following table presents a recapitulation of the result of the validity, reliability, power of difference, and difficulty level assessment instrument to produce report text for *Madrasah Aliyah* in two stages.

**Table 2.** Recapitulation of the validity, reliability, power of difference, and difficulty level assessment instrument

Assessment Area	Validity	Reliability	Power of Difference	Difficulty Level
Multiple Choice I	16 valid, 4 invalid	0,658	5 good, 6 enough, 9 bad	5 easy, 8 medium, 7 difficult
Multiple Choice II	20 valid, 0 invalid	0,755	2 excellent, 3 good, 6 enough, 9 bad.	10 easy, 6 medium, 4 difficult.
Short Answer I	8 valid, 2 invalid	0,443	3 excellent, 1 good, 1 enough, 5 bad.	1 medium, 9 easy
Short Answer II	10 valid, 0 invalid	0,528	3 good, 1 enough, 6 bad.	1 medium, 9 easy
Essay I	4 valid, 1 invalid	0,248	1 excellent, 4 enough.	5 medium
Essay II	5 valid, 0 invalid	0,325	1 good, 4 enough.	5 medium

Based on the validity and reliability test of the assessment instrument to produce report text, it can be

stated that the assessment instrument is effective and feasible to evaluate cognitive, skill, and attitude aspects in competence producing report text for the first-year

students of *madrasah aliyah*. This assessment instrument could support the curriculum of *Bahasa Indonesia* subject which is implementing text-based curriculum. This is because this assessment instrument is equipped with language, culture, literature, and communication objectives [27]. These findings correspond with Sheming Qu [28] affirming that the development of language intelligence includes linguistic, cooperative, cultural, and communicative.

#### 4. CONCLUSION

Based on the results, it can be concluded that the characteristics of the assessment instrument of producing report text for *Madrasah Aliyah* students consist of four objectives, language, cultural, literary, and communication objectives. The innovation of assessment instruments to produce report text for *Madrasah Aliyah* students consists of product design, teachers and expert evaluation of product design, and improvement of product design. The validity and reliability of the assessment instrument to produce the report text consist of validity and reliability of multiple-choice questions, short answer questions, and essay questions.

There are many implications of this research. For the teachers, they could apply this research product to improve the quality of learning outcomes in *madrasah aliyahs*. In addition, the innovation of this research can be used as a choice to evaluate the students learning outcomes, so that the teachers will be familiar with developing assessment instruments, enhancing teacher creativity, and creating creative learning activities with learners. For policymakers, this research could be used as a reference to prepare assessment instruments in Indonesian language subjects in all levels of school such as elementary, junior high school, and senior high school. The limitation of this research is that the research and development stage are incomplete so that, the further research might complete this research in large-scale trials.

#### AUTHORS' CONTRIBUTIONS

All authors conceived and designed this study. All authors contributed to the process of revising the manuscript, and at the end all author have approved the final version of this manuscript.

#### REFERENCES

- [1] Litbang Kemendikbud. Survey internasional PISA. accessed from <http://litbang.kemendikbud.go.id>. 2019. 21 September 2021.
- [2] P. Knapp, M. Watkins, Genre, text, grammar: Technologies for teaching and assessing writing, unsw Press, 2005.
- [3] D. R. Krathwohl, David R. Taxonomy of educational objectives: The classification of educational goals, Affective domain, 1964.
- [4] Kementerian Pendidikan dan Kebudayaan. Bahasa Indonesia wahana pengetahuan untuk SMA/MA kelas X. Jakarta: Kementerian Pendidikan dan Kebudayaan. 2013.
- [5] R. M Valette, Directions in Foreign Language Testing, 1969.
- [6] A. N. Khomsah, S. Subyantoro, Developing Evaluative Descriptive Text with Rebecca M, Valette's Taxonomy and CLIL Approach, Seloka: Jurnal Pendidikan Bahasa dan Sastra Indonesia 8(2), 2019, pp. 79-85.
- [7] O.P Wardani, Pengembangan perangkat evaluasi berdasarkan taksonomi the structure of observed learning outcome in LITERA Journal. Yogyakarta: Universitas Negeri Yogyakarta, Volume 13, Nomor 1, April 2014.
- [8] D. Coyle, CLIL: Planning tools for teachers, Nottingham: University of Nottingham, 2005.
- [9] P. Mehisto, Criteria for Producing CLIL Learning Material, Online Submission, 2012.
- [10] R. W. Brog, and M. D. Gall Educational research: an introduction. Fifth Edition: Longman. 1989.
- [11] I. D. G. P. Wirayasa, I. P. Darmayasa, Pengembangan Instrumen Penilaian Hasil Belajar Ranah Kognitif Model 4d Pada Materi Sepak Bola Berdasarkan Kurikulum 2013, Jurnal Pendidikan Jasmani, Olahraga dan Kesehatan Undiksha 8(3), 2021, pp. 81-88.
- [12] J. Klufa, Multiple choice question tests—advantages and disadvantages, 3rd International Conference on Education and Modern Educational Technologies (EMET), 2015.
- [13] N. Alruwais, G. Wills, M. Wald, Advantages and challenges of using e-assessment, International Journal of Information and Education Technology 8(1), 2018, pp. 34-37.
- [14] N. Pourdana, Exploring the Effects of Genre-based Portfolio Assessment on EFL Writing With Focus on Learner Engagement, 2021.
- [15] D. Kovacek, M. L. Bode, Authentic assessment for English language learners: Practical approaches for teachers, 1996.

- [16] I. Thirunavukarasu , The effects of cultural based text types in reading comprehension, *Journal of Nusantara Studies (JONUS)* 6(1), 2021, pp. 1-23.
- [17] A.R. Rauf, S. Sultana. Effect of faculty training on quality of Multiple Choice Questions. *Rawal Medical Journal* 46(2), 2021, pp. 430-433.
- [18] R. Misra, M. McKean, College students' academic stress and its relation to their anxiety, time management, and leisure satisfaction, *American journal of Health studies* 16(1), 2000, pp. 41.
- [19] J. B. Odeyemi, Critical Evaluation of Learning Duration and Students' Performance in Secondary School Mathematics, *International Journal of Women in Technical Education and Employment (IJOWITED)*, The Federal Polytechnic, Ilaro Chapter 2(1), 2021, pp. 29-34.
- [20] E.P Widoyoko, Penilaian hasil pembelajaran di sekolah, Yogyakarta: Pustaka Pelajar 1(2), 2014 pp. 8.
- [21] S.A. Karim, S. Sudiro, S. Syarifah, Utilizing test items analysis to examine the level of difficulty and discriminating power in a teacher-made test, *Utilizing test items analysis to examine the level of difficulty and discriminating power in a teacher-made test* 6(2), 2021, pp. 256-269.
- [22] J.M. Torres, L.M. Collantes, A.R. Millan, Classification of learning outcomes and assessment activities in CHED prototype and SUC syllabi based on Kratwohl's Taxonomy, *Elementary Education Online* 20(5), 2021, pp. 497-510.
- [23] J. W. Oller, *Language test at school*. New Mexico: Longman. 1979.
- [24] H. Yustanto , Effects of Prosody on The Politeness of Javanese Interrogative Sentences, *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 12(13), 2021, pp. 634-641.
- [25] E. P. Gracia, R. S. Rodríguez, A. P. Pedrajas, A. J. Carpio, Teachers' professional identity: validation of an assessment instrument for preservice teachers. *Heliyon* 7(9), 2021, pp. e08049.
- [26] M. Akdere, K. A. Clair, Y. Jiang, An examination of the effectiveness of virtual reality technology for intercultural competence development, *International Journal of Intercultural Relations* 82, 2021, pp. 109-120.
- [27] S. Qu, An Empirical Research on the Development of Students' Multiple Intelligences Based on the System of Diverse Assessment for College English Class, 3rd International Conference on Culture, Education and Economic Development of Modern Society (ICCESE 2019). Atlantis Press, 2019.