Analytical Rubric Development Design for Objective Test Assessment

Tuti Iriani[1*], Anisah[2], Yusrina Luthfiana[3], Johar Maknun[4], Nitih Indra Komala Dewi[5]

{tutiiriani@unj.ac.id[1], anisah_mt@unj.ac.id[2], luthfianayusrina@gmail.com[3], joharmaknun@upi.edu[4], nitih@upi.edu[5]}

Faculty of Engineering, Universitas Negeri Jakarta, Jakarta Timur, Indonesia[1,2,3],
Faculty of Technology and Vocational Education, Universitas Pendidikan Indonesia, Bandung, Indonesia[4,5]

Abstract. The goal of this study was to create an analytic rubric to evaluate students' abilities in compiling objective test assessments. This development research employs the Plomp development model, which is divided into five phases: (1) initial investigation, (2) design, (3) realization, (4) evaluation and revision, and (5) development. Instrumental trials were included in the Learning Evaluation course. This study's sample size is 115 students. The validity results show a measured value greater than 0.3, indicating that the instrument is valid. The correlation value between the criteria measured is greater than 0.5, indicating that all of the criteria measured can be factored by including all items, according to Kaiser Meyer Olkin's (KMO) 0.834 and Bartlett's anti-image correlation with a significance of 0.000. The seven criteria measured yielded two factors, according to exploratory factor analysis. Each criterion has a loading factor value greater than 0.3. The Alpha-Cronbach reliability test yielded a value of 0.800, and the instrument was declared reliable. The conclusion demonstrates that the analytic rubric for assessing performance ability in compiling objective questions is valid, reliable, and practical.

K**eywords:** Analytical rubric, Plomp model, objective test assessments

1 Introduction

The assessment determines the education system's progress and achievement levels for learning outcomes. However, good data sources are required, such as the measurement results. In general, measurement involves steps that assign numbers to the results of learning activities. The involved activities can be conducted through specifically designed tests according to the intended purpose, need, and preparation rules. Based on assessment in learning, Gardner [1] stated that assessment is divided into two forms. This includes both assessment of learning and assessment of learning for assessment.According to Pellegrino [2], content specification and the targeted assessment process provide the basis for developing initial assignments and rubrics that will provide evidence. Good assessment includes the assessment of learning and uses rubrics in the process. This is consistent with the findings of Rezaei et al. [3], who stated that the use of rubrics as a performance appraisal tool by teachers is becoming more popular. Through rubrics, teachers can be more objective in paying attention to assessments and obtaining other information to see student progress in completing their work. Rubrics are also useful in assessing learning that requires a high level of thinking, according to Drade et al. [4] and Wolf et al. [5]. In addition, students can also use rubric information to check the accuracy of assessments and increase their confidence in the fairness of their grades. Other benefits of rubrics for teachers include that they can be used in several

ACEIVE 2022, October 20, Medan, Indonesia

DOI 10.4108/eai.20-10-2022.2328882

learning contexts such as virtual learning, essays, practicums, presentations, exhibitions, and performances, portfolios of student work, artwork, and internships [6]. Rubrics also have many definitions and uses in education; for example, they can be used as scoring schemes [7] to guide people in assessing various constructs, such as the quality of student work, academic achievement, and educational resources [3][8][9]. According to Andrade [10] and Arter [11], the rubric contains assessment criteria that describe the level of quality. while Jönsson et al. [12] define rubrics as "assessment instruments designed to help identify and evaluate qualitative differences in student performance." In other words, a rubric is a rule, guide, criterion, or description designed to explain how to achieve a quality level. According to McGrath et al. [13], rubrics can be classified into holistic, analytic, specific, and general aspects. Holistic rubrics describe the overall quality of performance, while analytics assign multiple scores to the main features or aspects of an assignment. Customized rubrics have performance descriptions leading up to the assignment. while the General Rubric can assess all performance assignments for general aspects. Several recent studies have shown that the use of rubrics in education is very beneficial for students' learning and helps them achieve better results. In addition, it can help teachers assess the performance of their students. Rubrics also help students understand teacher expectations, clarify unclear goals, and ensure consistency of scores [5] [10] [14] [15]. According to Nkhoma et al. [16], several research findings have shown that rubrics are a useful assessment tool for students in exams and educational tools that support students in choosing the right learning approach, assist teachers in designing effective instructional strategies, and increase the reliability and validity of assessments. Muhammad et al. [17] conducted a holistic and analytical rubric analysis to assess Indonesian subjects. The results showed that an analytical rubric could not design a rubric consisting of 4 components, including criteria, weight performance level, and score components. However, the holistic rubric consisting of 4 data points was able to design a rubric with 2 components, specifically description and score. Mutiara et al. [18] developed a holistic and analytic scoring rubric using a tube material and applied the Plomp model. The holistic and analytic scoring rubrics were declared valid based on the percentage score (87.5%) based on the study results.

The purpose of this study was to describe the design and development of an analytical rubric used to assess students' abilities in preparing objective questions, specifically multiple-choice questions. Different from other assessment forms, multiple-choice tests can assess greater learning objectives; the focus is on reading and thinking skills based on standardized knowledge [19].

The analytic rubric used consisted of seven properties, specifically the preparation of question grids, objective questions, scoring, calculating the norm reference assessment and criteria reference assessment, calculating validity and reliability, calculating difficulty level, and adding discriminatory power to item analysis. The performance assessment of these students requires accuracy in calculation and analysis. The development of this rubric will improve the student performance assessment in compiling objective questions, specifically the results of the Learning Evaluation course. Currently, the teacher gives a direct score based on the work process and answers. The scoring description test becomes less objective, leading to a low student learning outcome. The effectiveness of learning activities is unclear, with inadequate information on the extent of students' ability to communicate problem-solving. Danielson et al. [20] stated that analytical rubrics help obtain sufficient information about students' strengths and weaknesses in communicating problem-solving results. This ensures analytic rubrics are suitable when distinguishing analytic levels. Guskey et al. [21] stated that the analytic rubric divides the product or performance into different dimensions and assesses

each separately so that the assessment becomes more detailed. A separate score is provided for each identified trait since an analytic rubric rates them independently. However, Levy [22] stated that problem assignments inform the teacher what the students know. According to Anil et al. [23], an analytical rubric is a scale used to assess students' ability to perform tasks for each criterion. Specifically, the analytic rubric states the performance level for each criterion to enable teachers to assess student performance. It requires the scorer to assess separate components or individual tasks related to the intended performance. The performance is scored separately, while each product part or performance is scored first and added to each score to obtain the total. The provision of the analytical rubric is the process of examining a person's work, requiring separate time for specific performance tasks or scoring criteria. The teacher pays attention to the assessment and obtains other information that monitors achievement progress and success for effective decision-making. Nkhoma et al. [16] state that clear criteria for evaluation and explicit quality definitions at given levels of achievement are key components of rubrics promoting students' learning. Students can effectively design their learning plans when informed of what is important. According to Atmazaki [24], the analytic rubric has 4 criteria, including 1 criteria, which is the first in designing the assessment rubric for the analytic component. (2) The weight, difficulty, and importance of performing a task correctly; (3) Performance, from the highest to the lowest level in the analytic rubric; and (4) Score, a number that shows students' performance. Therefore, the proposed rubric is expected to achieve an acceptable level of validity and reliability testing so that it can be used as an effective learning assessment tool. research and development of an analytic rubric to use the model [25]. Specifically, the procedures used in model development include the initial research, prototyping, and assessment stages.

2 Research Method

The research and development method is used in this study, which is a systematic approach to designing, developing, and assessing a program, process, or learning outcome in accordance with established consistency and excellence standards [26]. The Plomp model is used to create the rubric for the proposed model. According to the Plomp model, there are five stages: (1) preliminary investigation (2) Style (3) the realization/construction stage; (4) the testing, evaluation, and revision stage; and (5) the implementation stage. The following activities are described in each phase: (1) The preliminary investigation phase entails identifying and planning activities to define problems; (2) the design phase, which aims to design problem-solving, entails a systematic process to divide large problems into small problems; and (3) the realization or construction stage, which entails building a prototype with the main design based on the initial design document. (4) The quality of the learning design components to be developed should be considered during the test, evaluation, and revision stages. (5) The Implementation Stage, which is associated with the trial phase to validate the developed device. Figure 1 depicts the five steps in schematic form. Figure 1 depicts the five steps in schematic form.

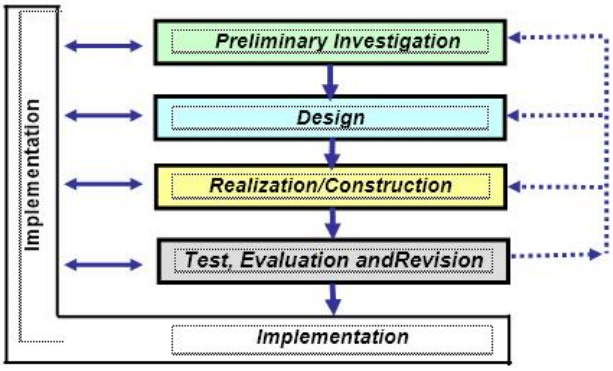

Figure 1. Model of Plomp

To determine the validity and reliability standards applied by selected experts from the first draft to the final version. Five experts were chosen as validators based on the following two criteria: (a) a lecturer in the course Evaluation of Learning; and (b) four validators with doctorates and one master's degree.

The rubric is assessed for validity using a 4-point Likert scale with seven (seven) indicators. A weight of 1 is given for each aspect and sub-aspect for scoring purposes. Likewise, there are four levels, each with a description of the performance. The compiled performance description refers to complete objective questions, preparation indicators, and assessment indicators. Specifically, this indicator conveyed accurate information in the event that the subject wrote it correctly and according to the steps for completing the right task. A weight of 1 is given for each aspect and sub-aspect for scoring purposes. Likewise, there are four levels, each with a description of the performance. The compiled performance description refers to complete objective questions, preparation indicators, and assessment indicators. Specifically, this indicator conveyed accurate information in the event that the subject wrote it correctly and according to the steps for completing the right task. The following shows the criteria for the validity of the rubric [27].

Table 1. Criteria for the validity of the rubric

Validity Criteria	Validity Level
85.01% - 100%	Very valid or can be used without revision
70.01% - 85.00%	Quite valid or can be used but needs a little revision
50.01% - 70.00%	Not valid, it is recommended not to use it because it needs a major revision
1.00% - 50.00%	Invalid or should not be used

To determine the effectiveness of the proposed model rubric, a validity test was conducted through factor analysis using the exploratory factor analysis method for the main components with the help of the SPSS 25 program.

3 Results and Discussion

This study was conducted in an attempt to produce an acceptable proposed model rubric for assessing student performance in compiling objective questions, and the results are presented herein.

3.1 Validity of the Proposed Rubric

The appropriateness of an assessment tool to measure what is intended to be measured is referred to as validity [28][29][30]31]. Determining the validity of an assessment tool is critical to ensuring that our efforts in using it are documented and that nothing is wasted. The assessment instrument for the Authentic Rubric is in the form of a performance assessment to measure students' ability in compiling objective questions validated by experts (expert judgment), namely 5 lecturers of Learning Evaluation courses from two universities, Indonesian Education University and Jakarta State University. In this study, validation includes the format, content, possible answers, and language. The goal of expert validation is to create valid instrument items in terms of content, material, and language, as well as quality products in terms of assessment format, rating scale, and assessment criteria. According to the experts' assessments, the validation results are as follows:

Table 2. Analysis Expert Judgement Data

No.	Aspect	Assessment Criteria	V1	V2	V3	V4	V5	Average	Average Aspect	%	Validity Level
1	Format	The rubric format is easy for users to understand	4	3	4	4	4	3.8	3.8	95	Very Valid
2	Language	The language used in the rubric is clear and not ambiguous	3	3	4	3	3	3.2	3.2	80	Quite Valid
3	Content	The aspects contained in the rubric provide solutions to the problem	3	4	4	4	3	3.6	3.6	90	Very Valid
		The description of each aspect encompasses all possible answer	4	4	3	3	2	3.2			
		The scoring scale corresponds to the results profession	4	4	4	4	4	4			
4	Possible Answer	The rubric provides all possible answer	4	4	3	4	3	3.6	3.6	90	Very Valid

Based on the results of the data analysis in Table 2, the average value of each aspect of the instrument validation assessment is as follows: 1) In terms of format, the average value is 3.8; 2) language terms get an average of 3.2; 3) in terms of content, an average of 3.6. And 4) an average of 3.6 was obtained in terms of possible answers. Thus, the average value of the total validity of the assessment instruments from the five validators is 3.55, or 89%. By

matching the total average with the validity category set out in Table 1, the assessment instrument developed by this researcher is included in the "very valid" category. Furthermore, based on the Coefficient of Content Validity of the Proposed Rubric, it shows that the Corrected Item Total Correlation in Table 3 is all above 0.3. As a result, aspects 1, 2, 3, 4, 5, 6, and 7 are declared valid. The content validity test in this study is based on the results of expert tests conducted by five validators. The results of the study show that the compatibility between the rates means that the instrument has relatively high stability [32]. This is in line with the study conducted by Rahmawan et al. [33] that concluded that performance appraisal is suitable to be used as a form of evaluation.

Table 3. Content Validity Coefficient of the Purposed Rubric

Indicator	Corrected Item-Total Correlation
Grid Settings	0.374
Preparation for Multiple Choice Questions	0.474
Calculating the Norm Reference Test and Criteria Reference Test	0.500
Calculating Validity	0.639
Calculating Reliability	0.594
Calculating Difficulty and Distinguishing	0.491
Item Analysis	0.664

3.1.1 Construct Validity Test Result

The resulting data is said to be feasible or can continue to meet the requirements for factor analysis if it meets the KMO MSA requirements of > 0.5 and the factor correlation value has a correlation coefficient of > 0.3 based on the data analyzed using the exploratory factor analysis approach. The KMO-M SA results are shown in Table 4 below.

Table 4. KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy	0.834
Bartlett's Test of Approx. Chi-Square	218.822
df	21
Sig.	0.000

It can be explained as follows based on the test results on the feasibility of using factor analysis to solve research problems: The Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy = 0.834, indicating that factor analysis was used correctly. The Bartlett test of sphericity of 21 with a significance of 0.000 < 0.05 indicates that factor analysis is suitable for data reduction.

Table 5. Validity Test Results

No.	Criteria Measured	Anti-Image Correlation	Information
1	Grid Settings	0.764	Valid
2	Preparation for Multiple Choice Questions	0.799	Valid

3	Calculating the Norm Reference Test and Criteria Reference Test	0.863	Valid
4	Calculating Validity	0.843	Valid
5	Calculating Reliability	0.861	Valid
6	Calculating Difficulty and Distinguishing	0.858	Valid
7	Item Analysis	0.821	Valid

The validity analysis results are visible in the anti-image matrix (MSA), particularly in the anti-image correction section, which is useful for determining which variables are suitable for use in factor analysis. Table 5 shows the results of the validity test.

Based on the anti-image column correlation, the correlation value of the criteria measured is greater than 0.5, indicating that all of the criteria measured are valid and factor analysis can proceed. The following step is to determine how many factors can be formed into a factor analysis with 115 samples after exploratory factor analysis using the SPSS version 25.0 program.

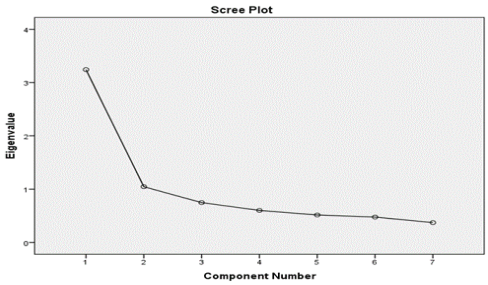

Figure 2. Scree Plot

According to the scatter plot above, three points appear to be greater than one, while the remaining points appear to be less than one. This demonstrates that there are two components with eigenvalues greater than one. There are two variables. The value of the loading factor > 0.3 indicates the grouping of the measured criteria and the magnitude of the loading factor of one of the factors. The Components Matrix Rotation Table, which displays the results of factor rotation, can be used to group items into factors. The grouping of the criteria measured into factors, as well as the magnitude of the loading factor obtained, can be seen in determining the input criteria measured against certain factors and the correlation between variables by factor, until a large correlation is obtained.

3.1.2 Rotated Component Matrix Test

Table 6. Rotated Component Matrix

	Component	
	1	2
Grid Settings	0.096	0.866
Preparation for multiple choice questions	0.269	0.778
Calculating the Norm Reference Test and Criteria Reference Test	0.724	0.039
Calculating Validity	0.748	0.253

Calculating Reliability	0.765	0.139
Item Analysis	0.802	0.197
Extraction Method: Principal Component Analysis		
Rotation Method: Varimax with Kaiser Normalization		

a. Rotation converged in 3 iterations

1) Aspects of the preparation of the grid. The correlation value with a factor of 1 = 0.096. and factor 2 = 0.866, because the relationship value of factor 2 > factor 1, the aspect of preparing the grid is included in factor group 2.
2) Aspects of the preparation of multiple-choice questions. The correlation value with factor 1 = 0.269 and factor 2 = 0.788, because the value of the relationship factor 2 > factor 1, the aspect of preparing multiple choice questions is included in the factor group 2.
3) Aspects of calculating the Norm Reference Test and Criteria Reference Test the correlation value with factor 1 = 0.724 and factor 2 = 0.039, because the relationship value of factor 1 > factor 2, the aspect of calculating the Norm Reference Assessment and Benchmark reference assessment is included in factor group 1.
4) Aspects of Calculation of Validity. The correlation value with factor 1 = 0.748 and factor 2 = 0.253, because the relationship value of factor 1 > factor 2, the aspect of arithmetic validity is included in the factor group 1.
5) Aspects of Calculating Reliability. The correlation value with factor 1 = 0.765 and factor 2 = 0.239, because the relationship value of factor 1 > factor 2, the Calculating Reliability aspect is included in the factor group 1.
6) Aspects of Calculation Analysis of the level of difficulty and distinguishing power. The correlation value with factor 1 = 0.606 and factor 2 = 0.222., because of the relationship value of factor 1 > factor 2, the calculation aspect of the level of difficulty and distinguishing power is included in the factor group of 1.50, so cannot be used to explain factors. Thus aspect 6 does not enter into factors 1 or 2.
7) Aspects of item analysis. The correlation value with factor 1 = 0.802 and factor 2 = 0.197, because the relationship value of factor 1 > factor 2, the aspect of the analysis item is included in the factor group 1.

The component transformation matrix shows that in component 1, the correlation value is 884 > 0.5 and in component 2, the correlation value is 0.884. Because the correlation value of all components is greater than 0.5, the two factors that form can be concluded to summarize the seven aspects analyzed. Referring to the interpretation above, it can be said that this factor analysis is as follows.

Table 7. Aspect Factor

Factor	Aspects
Factor 1	Grid Settings and Preparation of Objective Questions
Factor 2	Calculating Norm Reference Test and Criteria Reference Test, Calculate Validity, Calculation Reliability, and Item Analysis

Based on the aspect factors above, by reducing the results of the five variables, factor 1 is called planning, and factor 2 is called the result of implementation. The name is tailored to the stages that students must go through when evaluating their performance.

Furthermore, in the component plot in the rotatable space image, the seven variables are located on the two existing factors. Variables combined into one factor will be located close together at the same point. Meanwhile, a variable with a negative value will be located some distance away from other variables located at one point.

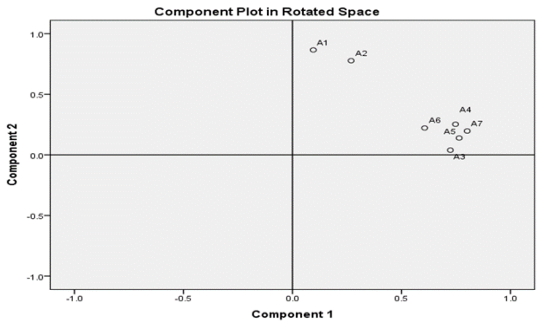

Figure 3. Component Plot in Rotated Space

3.2 Reliability of the Proposed Rubric

Following the completion of the construct validity test, the next step is to conduct a reliability test. The Cronbach Alpha reliability test was used in the Learning Evaluation course to assess instrument reliability using the SPSS 25.0 program. Based on the data processing results, the Cronbach's alpha value is 0.800, which is greater than 0.6. As a result, the questionnaire used in this study can be considered reliable. This study has a high level of trustworthiness. This is consistent with the theory, which states that if the reliability coefficient value is greater than 0.6, the expert is consistent in evaluating [34]. This means that the instrument is consistently evaluated by five experts. The researcher's analysis shows that the analytical rubric instrument developed is appropriate and meets the validity and reliability requirements for testing.

4 Conclussions

The development of the analytic rubric relies on the Plomp model, which involves three stages: initial research, prototype, and assessment. An analytical rubric is a set of performance scores for various independent evaluation criteria. It takes time to develop but provides accurate scores for both teachers and students. Based on the results of the confirmatory factor analysis, it can be concluded that all instruments are valid and can be trusted to provide information related to the ability to assess performance appraisals by preparing objective questions.

Acknowledgements

The authors express gratitude to Universitas Negeri Jakarta for the great support provided.

References

[1] J. Gardner. (2006). *Assessment for learning: A compelling conceptualization, Assessment and learning*, pp. 197–204.

[2] J. W. Pellegrino. (2002). Knowing what students know. *Issues Sci Technol*, vol. 19, no. 2, pp. 48–52

[3] A. R. Rezaei and M. Lovorn. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing writing*, vol. 15, no. 1, pp. 18–39.

[4] H. Andrade and Y. Du. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research, and Evaluation*, vol. 10, no. 1, pp.3.

[5] K. Wolf and E. Stevens. (2007). The role of rubrics in advancing and assessing student learning.," *Journal of Effective Teaching*, vol. 7, no. 1, pp. 3–14.

[6] S. Karkehabadi. (2013). Using rubrics to measure and enhance student performance.

[7] W. J. Popham. (1997). What's Wrong--and What's Right--with Rubrics. *Educational leadership*, vol. 55, no. 2, pp. 72–75.

[8] B. M. Moskal. (2000). Scoring rubrics: What, When and How. *Practical Assessment, Research, and Evaluation*, vol. 7, no. 1, pp. 3.

[9] D. Porcello and S. His. (2013). Crowdsourcing and curating online education resources. *Science (1979)*, vol. 341, no. 6143, pp. 240–241.

[10] H. G. Andrade. (2000). Using rubrics to promote thinking and learning. *Educational leadership*, vol. 57, no. 5, pp. 13–19.

[11] J. Arter. (2012). *Creating & Recognizing Quality Rubrics*. Portland: Pearson Assesment Training Institute

[12] A. Jönsson and E. Panadero. (2017). The use and design of rubrics to support assessment for learning in Scaling up assessment for learning in higher education*, Springer*, pp. 99–111.

[13] M. B. McGatha and P. Darcy 92010). Rubrics as play. *Mathematics Teaching in the Middle School*, vol. 15, no. 6, pp. 328–336.

[14] E. Panadero and A. Jonsson. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review, *Educ Res Rev*, vol. 9, pp. 129–144.

[15] T. H. Sundeen. (2014) Instructional rubrics: Effects of presentation options on writing quality. *Assessing Writing*, vol. 21, pp. 74–88.

[16] C. Nkhoma, M. Nkhoma, S. Thomas, and N. Q. Le. (2020). The role of rubrics in learning and implementation of authentic assessment: A Literature review in InSITE 2020*: Informing Science+ IT Education Conferences: Online*, pp. 237–276.

[17] K. Muhammad and M. Mukhlis. (2021). Analisis Rubrik Penilaian Biografi pada RPP Bahasa Indonesia. *J-LELC: Journal of Language Education, Linguistics, and Culture*, vol. 1, no. 1, pp. 27–36.

[18] A. D. Mutiara, A. Sutawidjaya, and A. Abadyo. (2017). Pengembangan Rubrik Penyekoran Pada Asesmen Otentik Untuk Materi Tabung," *Jurnal Pendidikan: Teori, Penelitian, dan Pengembangan*, vol. 2, no. 10, pp. 1393–1398.

[19] A. J. Nitko, *Educational assessment of students*. ERIC, 1996.

[20] C. Danielson and E. Marquez. (2016). *Performance Tasks and Rubrics for Middle School Mathematics: Meeting Rigorous Standards and Assessments*. Taylor & Francis [Online]. Available: https://books.google.co.id/books?id=yFypCwAAQBAJ

[21] T. R. Guskey and J. McTighe. (2016). Pre-assessment: Promises and cautions," *Educational Leadership*, vol. 73, no. 7, p. 38.

[22] M. Levy. (2004). Elementary and Middle School Mathematics: Teaching Developmentally. *Teach Child Math*, vol. 10, no. 5, pp. 286–288.

[23] Ö. Anil, M. Tatayoğlu, and M. Demirhan. (2012). Out-of-plane behavior of unreinforced masonry brick walls strengthened with CFRP strips. *Constr Build Mater*, vol. 35, pp. 614–624.
[1] J. Gardner. (2006). *Assessment for learning: A compelling conceptualization, Assessment and learnin*g, pp. 197–204.
[2] J. W. Pellegrino. (2002). Knowing what students know. *Issues Sci Technol*, vol. 19, no. 2, pp. 48–52
[3] A. R. Rezaei and M. Lovorn. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing writing*, vol. 15, no. 1, pp. 18–39.
[4] H. Andrade and Y. Du. (2005). Student perspectives on rubric-referenced assessment. Practical Assessment, *Research, and Evaluation*, vol. 10, no. 1, pp.3.
[5] K. Wolf and E. Stevens. (2007). The role of rubrics in advancing and assessing student learning.," *Journal of Effective Teaching*, vol. 7, no. 1, pp. 3–14.
[6] S. Karkehabadi. (2013). Using rubrics to measure and enhance student performance.
[7] W. J. Popham. (1997). What's Wrong--and What's Right--with Rubrics. *Educational leadership*, vol. 55, no. 2, pp. 72–75.
[8] B. M. Moskal. (2000). Scoring rubrics: What, When and How. *Practical Assessment, Research, and Evaluation*, vol. 7, no. 1, pp. 3.
[9] D. Porcello and S. His. (2013). Crowdsourcing and curating online education resources. *Science*, vol. 341, no. 6143, pp. 240–241.
[10] H. G. Andrade. (2000). Using rubrics to promote thinking and learning. *Educational leadership*, vol. 57, no. 5, pp. 13–19.
[11] J. Arter. (2012*). Creating & Recognizing Quality Rubrics.* Portland: Pearson Assesment Training Institute
[12] A. Jönsson and E. Panadero. (2017). The use and design of rubrics to support assessment for learning in Scaling up assessment for learning in higher education, *Springer*, pp. 99–111.
[13] M. B. McGatha and P. Darcy 92010). Rubrics as play. *Mathematics Teaching in the Middle School,* vol. 15, no. 6, pp. 328–336.
[14] E. Panadero and A. Jonsson. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review, *Educ Res Rev*, vol. 9, pp. 129–144.
[15] T. H. Sundeen. (2014) Instructional rubrics: Effects of presentation options on writing quality. *Assessing Writing*, vol. 21, pp. 74–88.
[16] C. Nkhoma, M. Nkhoma, S. Thomas, and N. Q. Le. (2020). The role of rubrics in learning and implementation of authentic assessment: A Literature review in InSITE 2020: *Informing Science+ IT Education Conferences*: Online, pp. 237–276.
[17] K. Muhammad and M. Mukhlis. (2021). Analisis Rubrik Penilaian Biografi pada RPP Bahasa Indonesia. *J-LELC: Journal of Language Education, Linguistics, and Culture*, vol. 1, no. 1, pp. 27–36.
[18] A. D. Mutiara, A. Sutawidjaya, and A. Abadyo. (2017). Pengembangan Rubrik Penyekoran Pada Asesmen Otentik Untuk Materi Tabung," *Jurnal Pendidikan: Teori, Penelitian, dan Pengembangan,* vol. 2, no. 10, pp. 1393–1398.
[19] A. J. Nitko. (1996). *Educational assessment of students*. ERIC.
[20] C. Danielson and E. Marquez. (2016). *Performance Tasks and Rubrics for Middle School Mathematics: Meeting Rigorous Standards and Assessments*. Taylor & Francis [Online]. Available: https://books.google.co.id/books?id=yFypCwAAQBAJ
[21] T. R. Guskey and J. McTighe. (2016). Pre-assessment: Promises and cautions," *Educational Leadership*, vol. 73, no. 7, p. 38.

[22] M. Levy. (2004). Elementary and Middle School Mathematics: Teaching Developmentally. *Teach Child Math*, vol. 10, no. 5, pp. 286–288.
[23] Ö. Anil, M. Tatayoğlu, and M. Demirhan. (2012). Out-of-plane behavior of unreinforced masonry brick walls strengthened with CFRP strips. *Constr Build Mater*, vol. 35, pp. 614–624.
[24] A. Atmazaki. (2013). Implementasi Kurikulum 2013 mata Pelajaran Bahasa Indonesia: Pola Pikir, Pendekatan Ilmiah, Teks (Genre), dan Penilaian Otentik. *International Conference on Languages and Arts*, pp. 15–22.
[25] T. Plomp et al. (2013). *Educational Design Research: An Introduction, no. pt. 1. Netherlands Institute for Curriculum Development* [Online]. Available: https://books.google.co.id/books?id=Zs7zoAEACAAJ
[26] R. C. Richey and J. D. Klein. (2014). *Design and development research: Methods, strategies, and issues.* Routledge.
[27] S. Akbar. (2013). *Instrumen perangkat pembelajaran*. Bandung: PT Remaja Rosdakarya.
[28] K. A. Korb. *Conducting Educational Research*," http://korbedpsych.com/ R09eValidity.html.
[29] D. Sigle "*Instrument Validity*. (2021). https://researchbasics.education.uconn.edu/ instrument_validity.
[30] G. M. Sullivan. (2011). A primer on the validity of assessment instruments. *Journal of graduate medical education,* vol. 3, no. 2.
[31] H. Taherdoost. (2016). *Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research. How to test the validation of a questionnaire/survey in a research*.
[32] N. K. Majid, T. J. Raharjo, and S. Supriyadi. (2017). Pengembangan Instrumen Asesmen Otentik Unjuk Kerja pada Mata Pelajaran IPA di SDN Jlamprang dan SDN Wonosari 03 Kabupaten Batang. *Journal of Research and Educational Research Evaluation*, vol. 6, no. 1, pp. 55–62.
[33] E. F. Rahmawan, T. Sumaryanto, and S. Supriyadi. (2016). Pengembangan instrumen penilaian kinerja kemampuan bernyanyi berbasis android. *Journal of Research and Educational Research Evaluation,* vol. 5, no. 1, pp. 81–89.
[34] S. Sujarwanto and A. Rusilowati. (2015). Pengembangan Instrumen Performance Assessment Berpendekatan Scientific Pada Tema Kalor Dan Perpindahannya. *Unnes Science Education Journal*, vol. 4, no. 1.