

The Analysis of Rubric Feasibility Using Video Snippets of Learning Process

Ghullam Hamdu^{✉1} & Iis Suryani²

^{1,2}Primary School Teacher Education Program, Universitas Pendidikan Indonesia, Tasikmalaya, Indonesia

✉ ghullamh2012@upi.edu

Abstract. Multiple interpretations can cause a decrease in the quality of rubrics. This research aims at developing an appropriate rubrics description for the measurement of students' scientific attitudes. The process of developing the rubrics was done by using the video snippets of learning process. A total of 23 observers whose final year students were involved in analyzing the video snippets of learning process. They assessed eight indicators of the scientific attitude of ten students through the video snippets. The feasibility of the rubric is considered based on the responses given by 85% of the observers who have the similar answer to the scientific attitude of students in the video snippets of learning process. The description test of the rubrics was analyzed descriptively, which then obtained two rubrics that obtained answer differences less than 85% of the observers. Therefore, this research focuses on investigating the two rubrics. This research implies that there needs to be a clear description of the rubrics in relations to the time of observation of the number of students and behaviors.

Keywords: feasibility analysis, learning process, rubrics, scientific attitude, video snippets

How to Cite: Hamdu, G., & Suryani, I. (2019). The Analysis of Rubric Feasibility Using Video Snippets of Learning Process. *Mimbar Sekolah Dasar*, 6(2), 239-252. doi:10.17509/mimbar-sd.v6i2.14150

INTRODUCTION~ Doing assessment through direct observation leads to subjective results. The subjective assessment results can be resolved if the indicators in the assessment are clear, and do not cause many behavioral interpretations of the object being assessed. One way to minimize subjective assessment is by developing a rubric as scoring reference to assess the achievement (Dornisch & Mcloughlin, 2006, p.1). A rubric is a descriptive scoring developed by teachers or other evaluators to guide the analysis of the product or process of students' effort (Moskal, 2000, p. 1). Scoring rubrics are often used to measure students' performance when assessing complex

performance-based tasks (Becker, 2016, p. 22). Scoring rubrics can be powerful tools for assessment and learning, if it is designed with the purpose and type of learning goal to be assessed in mind. (Ebert-may, 2010, 139).

Rubrics usually include a hierarchical score with a numerical scale, descriptive, or both of them, and each level has its own description (Goldberg, 2014, p. 1). Scoring rubric can be said to be good if the description of the scoring criteria is clear. Scoring criteria as an assessment tool must be valid and reliable. In fact, many teachers have known the terms 'validity' and 'reliability' in statistical studies, but they do not often discuss how these

concepts relate to classroom practice (Barbara & Leydens, 2000, p. 2). The discussion in this research is not presented statistically, but rather a descriptive study of the observers' responses in giving scores based on observations of students' behavior or performance shown in recording of video snippets from the learning process.

Rubric consists of several categories, hence, its developed description of the score in each category must be meaningful to clearly distinguish the performance. Each scoring category must be defined using the performance description rather than judgments about the performance (Moskal, 2000, p. 3). For example, the scoring rubric development sometimes occurs in a range of numbers using the limited description: Score 1 defines 'very poor', score 2 defines 'poor', score 3 defines 'good', and score 4 defines 'very good'. Assessment with this limited expression can lead to different kinds of perceptions. If this assessment method is used, it is very likely that the judgment of the assessors becomes more subjective. Perceptions for score 1 and 2, or 3 and 4 will be more subjective, since each score does not really define behavior differences. Therefore, the assessors are expected to explain different criteria. Regardless of how many levels of score are set, it is very important that these criteria can capture the definition of performance at each level (Goldberg, 2014, p. 2). A good rubric should be able

to eliminate the differences in the scoring of the same observation object at one time through a clear description.

Various methods are carried out by several researchers to analyze the use of the rubric as an assessment instrument. The development of rubrics to assess the learning outcomes in higher education conducted by Alsina et al. (2017, p. 151-155) consisted of several stages. First, describing the process of developing rubrics with their aspects, indicators, and levels. Second, presenting the results of the validation process through the assessment by external experts to confirm the significance of the instrument. Third, showing the results of the rubric application. They concluded that the new rubric produced could work well in assessing certain abilities. Its feasibility was also tested by external experts. This rubric result allowed different perceptions of the scoring rubric, if only a limited number of external experts are validated. Therefore, the validation of the scoring description of the rubric requires the use of a rubric to be tested by involving more field assessors extensively. The use of the rubric was tested with the aim of determining the equation of perception in interpreting the description of the intended rubric. The research conducted by Trace, Meier, & Janssen (2016, p. 32) showed that negotiations to determine mutual agreement in explaining ambiguous construction of the scoring rubric could produce more objective assessment of the

significant performance results. However, based on a study by Robin, Simon, & Robin, T. & Simon (2004, p. 3), the conclusions in the construction of the shared meaning still showed some errors in the rubric development. Among them are related to the use of language from the desired performance description. Joint negotiations in interpreting the description of the scoring rubric are not done, so the rubric can only be used and interpreted by the rubric makers.

This article discusses the rubric development by utilizing some video snippets of learning process. These video snippets were categorized based on students' behavior on the aspects and indicators that had been developed in the rubric. According to Jewitt (2012, p. 7), the selection of video snippets is an important part of analyzing videos and making explanations related to the use of criteria, and the purpose. Observation was done to analyze the video snippets. Observation is an important part of professional practice, and one of the most valuable and effective research methods (Caldwell & Atwal, 2005, p. 42). Observation, as a method of collecting research data, involves behavior in observing and systematically recording the results of these observations (Tobergte & Curtis, 2013, p. 1). In education context, video is used to observe the learning conducted by the teacher as a reflection material. Similarly, research conducted by Nagro et al. (2017, p. 15) showed that there was a

change in teachers' teaching behavior after seeing their own teaching video. Therefore, video recording in the learning process can also be considered as a reflection tool to improve the quality of learning.

This research seeks to provide recommendations on how to test the feasibility of a rubric and describe several rubrics with the acquisition of a level of similarity in the low observers' assessment. In determining the feasibility of the rubric, the observers are asked to assess some behavioral attitudes of students through the video snippets of the learning process.

Reference to aspects of scientific attitudes in question is based on studies of social attitude competencies from curriculum documents applicable in Indonesia. The selection of aspects of scientific attitudes in this research was determined by referring to the existed aspects of attitude in the curriculum, which was adjusted to the learning steps that had been implemented.

METHOD

The mechanism of the implementation of the feasibility test of the rubric was carried out qualitatively and descriptively through the responses given by the observers. The general stages of the feasibility test for this rubric consisted of three steps, as follows:

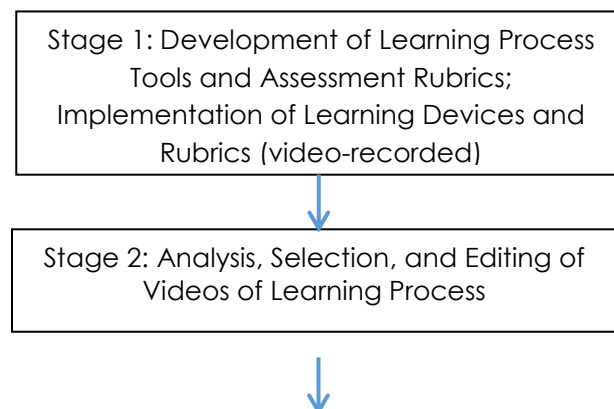
First, developing lesson plan. Based on the analysis of the learning process that had been made, several aspects and rubrics

were prepared to be tested. The rubric had eight scientific attitude indicators. The lesson plan was implemented in the classroom, then the learning process was video-recorded.

Second, sorting necessary videos in accordance with the aspects and rubrics that had been developed. The selection of the behavioral scene of students' scientific attitudes was taken from the total video during the learning process. The video footage could be considered as a substitute for direct observation of students' behavior in the classroom. The feasibility test was done based on several indicators of scientific attitude that have been developed previously. After obtaining the required video snippets, several observers were involved in testing the assessment rubric. They consisted of 23 college students in their senior year, consisting of 10 male students and 13 female students in the 2018/2019 Academic Year. They were asked to observe the eight behaviors of 10 five-

graders (students aged between 10-11 years) in the video snippets. Observations were conducted to students' behaviors related to scientific attitude indicators in the rubrics tested.

Third, analyzing the results of the observers' answers based on scientific attitude indicator in the rubric observation sheet. The results were converted into percentage. If the average percentage score showed similarity of less than 85% of the total observers, then the revisions made were on words and sentences construction to be re-tested. On the other hand, if the average percentage score showed similarity of more than 85% of the total observers, the rubric was considered feasible, and could be used extensively. The assessment of similarity score refers to the number of similarities in giving behavior scores to students' scientific attitude indicators by a number of observers. The following figure depicts the feasibility test of the rubric:



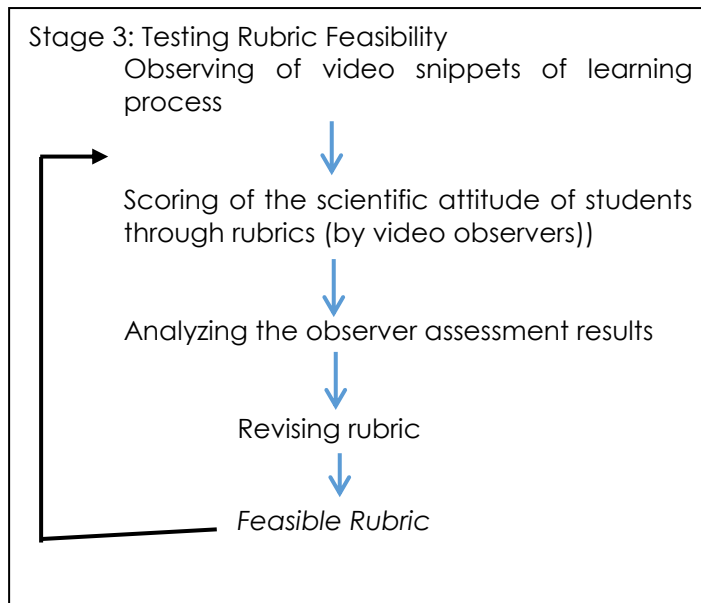


Figure 1. Flow of Feasibility Test through Analysis of the Learning Process Videos.

The focus of the following discussion is about the analysis of the feasibility test of the rubric (Stage 3), since this stage was the main activity in testing the rubric feasibility. Stage 1 and Stage 2 were not discussed since they were the initial procedures and required a separated discussion.

RESULTS AND DISCUSSIONS

The feasibility test of the rubric focused on several aspects of scientific attitudes in during the learning process done by individual student, as follows:

- 1.a: Caring for the environment attitude, with the indicator of maintaining the cleanliness of the school environment;
- 2.a: Curiosity attitude, with the indicator of enthusiastically answering teacher

questions;

2.b: Curiosity attitude, with the indicator of paying attention to the observed object;

3.a: Open-minded attitude, with the indicator of respecting other people's opinions or findings;

4.a: Critical thinking attitude, with the indicator of asking about any new changes or things;

4.b: Critical thinking attitude, with the indicator of repeating activities carried out;

5.a: Diligent attitude, with the indicator of working diligently on instructed tasks;

6.a: Careful attitude, with the indicator of re-checking answers



Figure 1. Scene of Learning Video Snippets from the Attitude of 1.a (Caring for the Environment Attitude).

After obtaining the video on the learning process, the video was processed by selecting scenes related aspects and indicators of the scientific attitude being tested.

Field Test 1

The implementation of Field Test 1 was carried out by observing the same 10 students. The same observations of each student were carried out by the observers

together. The videos were played one by one, then the observers responded to the students' behavior, and assessed them by giving a score based on the aspects and indicators in the observation sheet. The results obtained the observers' similarity score, the ideal score and the consensus results from 23 observers. The scores given by the observers to the 10 students are presented in the following Table 1 and Table 2.

Table 1. Results of the Assessment from the Observers to the Students 1 to 5 in Test 1.

Aspect	Student 1				Similarity (%)	Student 2				Similarity (%)	Student 3				Similarity (%)	Student 4				Similarity (%)	Student 5				Similarity (%)
	1	2	3	4		1	2	3	4		1	2	3	4		1	2	3	4		1	2	3	4	
1a		2		1	95		2			100		2			100	1	2			95	1	5			78
2a			1	2	95	2	1	5	1	65	9	1	1		56	1	1	3		78		2	1	2	86
2b		2			100		2			100		2			100	2				100	2				100
3a	2	4	6	1	47	1	1	2	6	60		8	5	1	43	1		5	1	73		2		3	86
4a			2	1	95	2		1		95	2				100	2				100	2				100
4b		2			100		2			100		2			100		2			100		2	2	3	86
5a		1	2		91	1		1	2	95		5	1	1	73			1	2	95			1	2	95
6a		1	2		95			2	2	91		2	2		91			2	2	91			2	2	91

Table 2. Results of the Assessment from the Observers to the Students 6 to 10 in Test 1.

Aspect	Student 6				Similarity (%)	Student 7				Similarity (%)	Student 8				Similarity (%)	Student 9				Similarity (%)	Student 10				Similarity (%)	
	1	2	3	4		1	2	3	4		1	2	3	4		1	2	3	4		1	2	3	4		
1a	2	2	1		95	2	2	1		95	2	1	2		95	1	2	2		95		2	2	1		95
2a		6	1	7	43		3	9	1	47		9	6	8	39	1	6	8	8	34		2	1	1		47
2b		2	3		100		2	3		100		2	3		100	1	2	2		95		2	3			100
3a	1		5	1	73		1	3	1	82	1		6	1	69	8		5	1	43	3	1	8	2		78
4a	1	2	2		95	2	3			100	2	2		1	100	2	3			100	2	3				100
4b				2	100				2	100				2	100				2	100				2	100	
5a				2	100			2	2	91		3	1	6	4	69			2	100		3		2	86	
6a			2	2	91		1	3	1	82			1	9	4	82		1	2	95		2	1	2	86	

Table 1 and Table 2 show the observers' assessment of students' scientific attitudes in the video snippets of the learning process. The results were obtained from the observation sheet in the rubric based on the observed aspects. The observers' assessment was then calculated based on the assessment given (1, 2, 3, or 4). The results were then converted into percentage according to the number of observers who had chosen 1, 2, 3, or 4 of the assessments per aspect for each student in the video.

After calculating the average percentage

of the 10 students, the average similarity score of responses from each aspect was obtained. The results of the average percentage for each aspect and the feasibility of the rubric based on Test 1 results are presented in Table 3.

Table 3 shows that there were two rubrics (2a and 3a) that were not feasible because the average percentage of observers' assessments was not yet suitable, which was below 85% of the total observers. Thus, those rubrics needed to be revised and re-tested.

Table 3. Average Percentage for Each Aspect and the Feasibility of the Rubric in Test 1.

Aspects	Average Percentage	Rubric Feasibility
1a	94.3%	Feasible
2a	59%	Revision
2b	99.5%	Feasible
3a	65.4%	Revision
4a	98.5%	Feasible
4b	98.6%	Feasible
5a	89.5%	Feasible
6a	89.5%	Feasible

Revisions were made by making the rubric description more specific. The substitution of words and sentences was done by considering language that might create multiple interpretations of the behavior displayed by students on the related to the number of behaviors shown by students and a simple description of the perception to an object/event experienced by students.

Discussion of the revised form of the rubric in aspect 2a (curiosity attitude with the indicator of enthusiastically answering

teacher questions).

A rubric with aspect 2a (curiosity attitude with the indicator of enthusiastically answering teacher questions) showed that the changes occurred were related to the elimination of the number of students answering teacher questions. The data showed that the rubric was not feasible to use because the percentages of the observers; assessment of similarity was still below 85%. After analyzing the rubrics and video snippets, it turned out that the video shown did not match the desired criteria.

Table 4. Revision of Rubric in Aspect 2a (Curiosity Attitude with the Indicator of Enthusiastically Answering Teacher Questions).

Aspect	Indicator	Before Revision	After Revision
2	a	<p>4 = The student answers all the questions asked by the teacher.</p> <p>3 = The student answers 3 questions from the questions asked by the teacher.</p> <p>2 = The student answers 2-1 questions from the questions asked by the teacher.</p> <p>1 = The student does not answer the questions asked by the teacher</p>	<p>4 = The student answers questions on their own initiative.</p> <p>3 = The student answers questions because the student is appointed by their friends.</p> <p>2 = The student answers questions because the student is appointed by the teacher.</p> <p>1 = The student does not want to answer questions.</p>

The number of teacher questions answered by students was considered, therefore, the observers provided a diverse rubric assessment of students' behavior in the video. It was suspected that the video did not show the number of students answering questions. However, the actual learning process was difficult to observe, since the observers took longer to observe the behavior of 10 students in the video

snippets, and to calculate the number of answers each student in learning. Furthermore, it was difficult to get a question and answer between the teacher and students with a certain amount (in the rubric of more than three answers) in one learning frame with limited observation time. Further analysis, if the rubric mentioned number of students answering teacher questions, it was necessary to

provide a number of video snippets that were likely to be long, related to each observed student answering some questions. This was not done due to considering the time, boredom, and focus of the observers on observing a number of video snippets with the duration of time and the number of videos. Therefore, in

order to overcome this problem, the description of the rubric was changed to be more about the initiative or encouragement (own initiative, encouraged by the teacher, encouraged by other students, or not answering questions).



Figure 2. Scenes of Learning Video Snippets from Attitude 2.a (Curiosity Attitude with the Indicator of Answering Teacher Questions Enthusiastically).

Discussion of the revised form of the rubric in aspect 3a (open-minded attitude with the indicator of respecting other people's opinions or findings)

other people's opinions or findings) indicated a revision related to the perceptions of observers on students' behavior in video snippets for the condition of students showing an attitude of attention to the opinions of others.

The rubric with aspect 3a (open-minded attitude with the indicator of respecting

Table 5. Rubric Revision on the Aspect 3 Indicator a (Open-minded Attitude with the Indicator of Respecting Other People's Opinions or Findings).

Aspect	Indicator	Before Revision	After Revision
3	a	<p>4 = The student listens to their friends' opinions and does not make noise when their friends express their opinions.</p> <p>3 = The student listens to their friends' opinions, but occasionally makes noise when their friends express their opinions.</p> <p>2 = The student listens to the opinions of his friends, but laughs</p>	<p>4 = The student pays attention to the opinions of their friends from beginning to end without any disturbing behavior</p> <p>3 = The student is not paying attention to the opinions of their friends, occasionally chatting or playing objects around.</p> <p>2 = The student pays less attention to the opinions of their</p>

<p>when the opinions of their friends are wrong.</p> <p>1 = The student does not listen to their friends' opinions.</p>	<p>friends, frequently chatting or playing objects that are around.</p> <p>1 = The student does not pay attention to the opinions of his friends from beginning to end.</p>
---	---

The statements that contained dualism in different circumstances were changed. For example, "listens to opinions, but..." (In rubric 2 and 3 before revision). The statement was seen uncertain and can be used in different situations. It was considered that the use of language with positive intentions followed by negative statements could lead to multiple interpretations. Then, there was a need for consistency in the use of words and sentence language between the situations clarified by another sentence to reinforce the situation. Therefore, several other similar situations were used but distinguished in students' behavior frequency, which was shown to clarify previous negative statements, such as sentence statements for rubrics 2 and 3 (the revised rubrics).

3 = the student did not pay attention to the opinions of their friends, **occasionally** chatting or playing objects around.

2 = the student paid less attention to the opinions of their friends, **frequently** chatting or playing objects around.

The analysis of changes in the rubric was taken into account since stating clear situation mutually reinforcing between the statement and the rubric was important. The developed rubric must consider the use of methods that was easy to understand so that they could be used to specifically assess the students' behavior through observation.

Similar to the revised form of rubric 2a, the frequency or number of these behaviors in the revision of section 3a could be observed directly at one time. The observers paid attention to one frame of the video that contained the students' behavior when expressing their opinion. This situation in the video could certainly be well observed without taking a long time.



Figure 2. Scenes of Learning Video Snippets from Attitude 3.a (Open-minded Attitude with the Indicator of Respecting Other People's Opinions or Findings).

Field Test 2

The second field test was conducted on the revised rubric, especially on aspects 2a and 3a. There were 20 video snippets observed (consisting of each of 10 videos for each observed student with 2 observed

aspects). The current observers were different from the observers in Field Test 1, but the criteria used were the same. This was done so that the results obtained were more authentic by assuming this rubric could be used by anyone later.

Table 6. Assessment Results from the Observers to the Students 1 to 5 in Test 2.

Aspect	Student 1				Similarity (%)	Student 2				Similarity (%)	Student 3				Similarity (%)	Student 4				Similarity (%)	Student 5				Similarity (%)
	1	2	3	4		1	2	3	4		1	2	3	4		1	2	3	4		1	2	3	4	
2a				2	100	2				100	2				100	2				100	2				100
3a				2	100	2		2		91				2	100			2		100			2		100

Table 7. Assessment Results from the Observers to the Students 6 to 10 in Test 2.

Aspect	Student 6				Similarity (%)	Student 7				Similarity (%)	Student 8				Similarity (%)	Student 9				Similarity (%)	Student 10				Similarity (%)
	1	2	3	4		1	2	3	4		1	2	3	4		1	2	3	4		1	2	3	4	
2a	2				100	2				100	2				100	2				100	2				100
3a	2	2			91	2	2			91	2	2			91				2	100			2		100

The results of the second test showed the overall assessment of the observers to the 10 students on aspects 2a and 3a. In

general, the observer's assessment of each student in the video snippets showed a great similarity, namely above 85%.

Table 8. Average Percentage for Aspects 2a and 3a and the Feasibility of the Rubric in Test 2.

Aspect	Average Percentage	Rubric Feasibility
2a	100%	Feasible
3a	96.4%	Feasible

The average results of the second test also showed that aspects 2a and 3a had a mean of 100% and 96.4% respectively, so that the rubric could be said to be feasible.

Implications of developing Rubrics with the help of video snippets of the learning process

The assessment tool was considered good if it could measure what should be measured, not affected biased measurement results, and could justify the results of these measurements. Measurements using observation sheets were different compared to using ordinary written tests. Measurements by observations focused more on the behavior of students directly at that time, while the measurement by of written test results could be done at certain times by checking the results of the answers. Assessment of students' behavior with direct observation requires clear and directed instruments, and also the preparation on what aspects of behavior were observed at the time of learning. The more important thing was whether the learning carried out by the teacher brought changes in students' behavior. Therefore, the intention of the assessments was to improve the quality of learning. Assessments using an observation tool must clearly state the measured aspects and indicators. Rubrics to describe the achievement of competence must also be clear in size.

A rubric had not been considered good if it was still interpreted differently by some assessors/observers (Wulan, 2018, p. 52). In the development of the rubric, negotiation to equalize perception was one of the effective methods to guide scoring and reduce the tendency of assessors to be subjective. Negotiation was a good way for assessors to build together interpretations of the construction of the language being measured (Trace et al., 2016, p. 42). The development of scoring rubric descriptions with the help of video snippets of the learning process had provided an alternative of the feasibility test for the rubric. There had been many research examining the use of videos in learning, but not focusing more on helping students to understand learning (Most, Academy, Selin, & Education, 2015, p. 3). Therefore, this research seeks to examine the use of video as an important instrument in helping to validate the development of a scoring rubric.

The video in question came from the recording of the learning that had been cropped into shorter videos. The video-cropping process was adjusted to the description of the scoring rubric being tested. By using the video snippets, the observers were not required to be present directly in the learning. The observers could simply watch a certain part of learning. This method was very practical because it could involve more observers in obtaining the better validity and feasibility of using the scoring rubric. The main

requirement that must be fulfilled by scoring developer is to record the learning process or choose certain learning activities and then check them accordingly with aspects and indicators to be measured and developed.

CONCLUSIONS

This research had provided a way to test the feasibility of a scoring rubric. Tests were carried out by analyzing the results of the scores given by the observers to the students' behavior displayed on the video snippets of the learning process. The feasibility of a rubric was determined by the level of similarity of the assessment results in each observer rubric. Clarity in describing the level of students' behavior was important to be considered because it was related to the observers' perception of what they had observed. Descriptions of behavior attainment had to be specific so that it did not cause unclear interpretations. The essence of this rubric test was to analyze the similarity of the scoring given by the observer to the students' behavior through the description of the assessment in the rubric. The more similarities in the scores given by the observer to the students' behavior mean the better the description of the behavior in the rubric. The video snippets of the learning process were only used as a tool to ease rubric developers to test the feasibility of scoring rubric descriptions. However, the results of this research revealed that the use of scoring descriptions based on the frequency of

emergent behavior could be incorrectly used when observing the videos based on the frequency of the desired behavior, which was difficult to calculate. In addition, the wording was also important to avoid multiple interpretations. The use of words or sentences in a category (score) in description had to support each other and be clear.

REFERENCES

- Alsina, Á., Ayllón, S., Colomer, J., Fernández-Peña, R., Fullana, J., Pallisera, M., ... Serra, L. (2017). Improving and evaluating reflective narratives: A rubric for higher education students. *Teaching and Teacher Education*, 63, 148–158. <https://doi.org/10.1016/j.tate.2016.12.015>
- Barbara, M., & Leydens, J. A. (2000). Scoring Rubric Development: Validity and Reliability. - *Practical Assessment, Research & Evaluation*, 7(10), 1–6.
- Becker, A. (2016). Student-generated scoring rubrics: Examining their formative value for improving ESL students' writing performance. *Assessing Writing*, 29, 15–24. <https://doi.org/10.1016/j.asw.2016.05.002>
- Caldwell, K., & Atwal, A. (2005). Non-participant observation: using video tapes to collect data in nursing research. *Nurse Researcher*, 13(2), 42–54. <https://doi.org/10.7748/nr2005.10.13.2>

42.c5967

- Dornisch, M. M., & Mcloughlin, A. S. (2006). Limitations of web-based rubric resources: Addressing the challenges. *Practical Assessment Research Evaluation*, 11(3), 1–8. <https://doi.org/http://dx.doi.org/Article>
- Ebert-may, D. (2010). *Scoring Rubrics. Educational Testing Service, Portland, OR, USA*, 1–17. Retrieved from [http://www.peopledev.co.za/library/Scoring rubrics - Moskal B.pdf](http://www.peopledev.co.za/library/Scoring%20rubrics%20-%20Moskal%20B.pdf)
- Goldberg, G. L. (2014). Revising an Engineering Design Rubric: A Case Study Illustrating Principles and Practices to Ensure Technical Quality of Rubrics. *Practical Assessment, Research & Evaluation (PARE)*, 19(8), 7714. Retrieved from <http://pareonline.net/getvn.asp?v=19&n=8>
- Jewitt, C. (2012). An introduction to using video for research An Introduction to Using Video for Research. *National Centre for Research Methods Working Paper*, 1–22.
- Moskal, B. M. (2000). Scoring Rubrics : What , When and How ? *Practical Assessment Research & Evaluation*, 7(3), 3–7.
- Most, I., Academy, K., Selin, J., & Education, H. (2015). Video - Based Learning: Current Research on Benefits and Best Practices, 1–8.
- Nagro, S. A., deBettencourt, L. U., Rosenberg, M. S., Carran, D. T., & Weiss, M. P. (2017). The Effects of Guided Video Analysis on Teacher Candidates' Reflective Ability and Instructional Skills. *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children*, 40(1), 7–25. <https://doi.org/10.1177/0888406416680469>
- Robin, T., Simon, M., & Robin, T. & Simon, M. (2004). What's still wrong with rubrics: focusing on the consistency of performance criteria across scale levels. *Practical Assessment, Research & Evaluation*, 9(2), 1–6. Retrieved from http://www.asu.edu/courses/asu101/asuonline/temp/whats_still_wrong_wit_h_rubrics.pdf
- Tobergte, D. R., & Curtis, S. (2013). Observation in Research. *Journal of Chemical Information and Modeling*, 53(9), 1689–1699. <https://doi.org/10.1017/CBO9781107415324.004>
- Trace, J., Meier, V., & Janssen, G. (2016). "I can see that": Developing shared rubric category interpretations through score negotiation. *Assessing Writing*, 30, 32–43. <https://doi.org/10.1016/j.asw.2016.08.001>
- Wulan, AR (2018). *Menggunakan Asesmen Kinerja untuk Pembelajaran Sains dan Penelitian*. Bandung: UPI Press.