

A Rubric Study for Assessing Paragraph Level Written Texts

Şebnem İlhan Agan¹, Sabahattin Deniz²

¹Muğla Sıtkı Koçman University, School of Foreign Languages, Turkey

²İzmir Demokrasi University, Faculty of Education, Turkey

Correspondence: Şebnem İlhan Agan, Muğla Sıtkı Koçman University, School of Foreign Languages, Turkey.

Received: November 11, 2019

Accepted: December 5, 2019

Online Published: December 16, 2019

doi:10.11114/jets.v8i1.4610

URL: <https://doi.org/10.11114/jets.v8i1.4610>

Abstract

In this study, an analytical rubric which has a five dimensional scoring key was developed. It was used to evaluate paragraph-level written texts which were written by the students attending process writing activities in B1 optional preparatory classes in Muğla Sıtkı Koçman University School of Foreign Languages. The development stages of rubric are as follows: establishing the dimensions of the rubric, establishing the performance levels, cells (the criteria of the rubric) and the performance level score ranges of the rubric, examining the usability of the rubric by forming sample texts, the trial of the rubric, the validity and the reliability of the rubric. In the development stages of the rubric; for the pilot study, 60 university students in the optional B1 English preparatory program at Muğla Sıtkı Koçman School of Foreign Languages and an English language lecturer were included. Also, the opinions of three English language lecturers' and two experts' who are in the field of education sciences and one instructor's who is in the field of Turkish language and literature were taken, to estimate the inter-rater reliability two English language lecturers were included. When Kappa coefficients of the dimensions of the rubric were examined, it can be said that there was a moderate and strong agreement among the raters in terms of assessing all dimensions in the rubric.

Keywords: writing rubric, analytical rubric, developing a rubric, inter-rater reliability

1. Introduction

Foreign language learning in Turkey is gaining more importance due to the global developments in the world and is included in school programs at the primary school education level. As a foreign language, English ranks first among the foreign languages aimed to be learned. Foreign language learning is an expensive process starting from the second year of primary education and lasting more than a decade, including university. Nevertheless, it is clear that the desired goals in foreign language teaching have not been achieved. This is mostly due to the use of traditional methods and techniques in foreign language teaching (Gömleksiz & Elaldı, 2011).

It is stated that writing skill, which is among the basic language teaching skills in foreign and native language teaching, is seen as the most difficult skill in terms of learning and comprehension (Onozawa, 2010; Graham & Sandbell, 2011). Writing skill is seen as an area where students do not feel completely comfortable and secure because it is a skill that requires full and correct expression. Moreover, for writing skill to be performed completely and correctly, it is important for the writer to use grammar and spelling-punctuation correctly and properly, and at the same time to render the text satisfactory and accurate in terms of word selection, form and content (Selvikavak, 2006).

Bayram (2006) emphasizes that many foreign language learners regard writing skill as a difficult process and expresses that learners must consider various factors before, during and after the writing process. When it comes to academic writing, the burden on the learner increases further, as he/she has to learn and apply academic writing while learning the target language. The fact that they cannot see their progress while trying to write may lead them to despair and giving up from time to time.

According to Collins and Gertner (1999, cited by Inan, 2014); the fact that these four basic structural levels, including the full textual structure, paragraph structure, sentence structure (morphology) and word structure, cannot be formed correctly in the native, second or foreign language at the same time, causing writing skills to be perceived as difficult. Moreover, the fact that the importance given to implementation, evaluation and correction in classroom activities is insufficient, and that although feedback is given, it is mostly aimed at correction or classification delays the acquisition and development of this skill (Chastain, 1988).

In this day and age, when, in addition to basic knowledge and skills, it is necessary to provide students with high level knowledge and skills such as analytical thinking, critical thinking, creative thinking and problem solving, it must be correctly determined whether the projected knowledge, skills, attitude and behavior is acquired or not and to what extent if acquired (Atmaz, 2009).

In this context, assessment and evaluation is an inseparable and supplementary part of the education process. Through assessment and evaluation practices, teachers can evaluate the effectiveness of the program and teaching, and identify students' learning deficiencies in order to determine their success (Stark, 1998).

Commonly used techniques such as written, multiple choice, short-answer techniques to measure student achievement rather focus on students' exam skills, exam results and rankings derived from said results, and ignore students' thinking skills. Considering that student performance changes in different situations and times, it can be said that these techniques, which are performed in a short period of time and assess student success by looking at their positions within the group, are inadequate, especially in evaluating the process. The student-centered teaching approach expressed in the curricula steers the teacher away from being an authority center into a guiding figure. It is essential to educate individuals who can access and use information with the active participation of the student in the learning environment. The aim of the curricula is to reveal the interests and abilities of the students by following them in the learning process and to ensure that the process is evaluated in addition to the product. This case has brought new assessment techniques with it (Parlak & Doğan, 2014).

As a result of the assessment process, certain decisions are made about certain subjects related to individuals. The accuracy of the decisions depends on whether the measurement on which the assessment is based has no or as few as possible errors, gives consistent results in repeated measurements and is related to the decision to be made. In other words, it is necessary to ensure the reliability and validity of the measurement tools to be used (Atmaz, 2009).

In practice, when an object or attribute is measured at different times, it is impossible to achieve the same result in all of the measurements. Because, for various reasons, a measurement that is completely free from error cannot be made. Therefore, individual scores obtained from one test will vary from one measurement to another. Errors involved in measurement results may arise from the measurer, the measurement tool and the measurement method, the environment in which the measurement is made, and the interaction of the measured individuals with these factors (Tekin, 2004).

In this context, rubrics appear to be the most preferred measurement tools today. Rubrics are a guideline of the basic characteristics of a skill, as well as gradually assessing a skill as a product or performance (Beyreli and Konuk, 2016). There are many different definitions of rubric in the literature;

According to Haladyna (1997), rubric is an assessment tool that allows us to record the scores of our observations to appropriate dimensions in the defined categories (criteria or criteria). According to Popham (1997), it is a scoring tool that lists the criteria for each study and shows what to do in the study. Callison (2000) defines it as the set of criteria determined between the unacceptable lowest level and the highest observable level of performance. Moskal (2000) defines it as a scoring design used to analyze the products and performances of individuals in detail. According to Sezer (2005), rubric is the assessment of the expected student performance at the end of a certain teaching process by dividing it into different dimensions and levels.

Rubrics are divided into two main categories as holistic and analytical. Holistic scoring is preferred because it is assumed to be easy, cheap and accurate when the observer is required to make a general decision on performance quality in a large-scale assessment (Jonsson & Svingby, 2007). Holistic rubric requires assessing the performance or the resulting product as a whole without judging the components of the performance or the resulting product (Moskal, 2000).

There are some cases where holistic rubric is preferred. If the performance to be measured cannot be divided into parts and there is overlap between the criteria for performance, if the resulting product or performance is excellent, or if errors in certain process and procedure steps can be ignored. It is preferred if the student is expected to have an answer or solution reflecting his/her own learning style and there is no certain and definite answer, if a holistic score regarding student performance is desired, or if the student's studies will be assessed in a short time (Moskal, 2000; Mertler, 2001; Nitko, 2001; Kan, 2007).

In analytical rubrics, the evaluator assigns a score for each of the dimensions assessed in the task. It is useful in the classroom because the results can help teachers and students to identify the strengths and learning needs of students (Jonsson & Svingby, 2007).

And In some cases analytical rubric is preferred are; If it is necessary to assess the components of the performance or the resulting product separately and independently in line with the main criteria, if the performance to be measured is multidimensional and can be divided into components, if it is wished to provide feedback to teacher and student

regarding the performance task, or if there is enough time for assessing the performance (Moskal, 2000; Mertler, 2001; Nitko, 2001; Kan, 2007).

Rubrics can be used by both the teacher and the learner. The use of the expressed measurement tool by the learner helps the learner gain self-awareness and develop self-regulation skills about the performance wished to be measured or assessed (Melanlioğlu, 2016). Saddler and Andrade (2004) emphasize this point by stating that rubrics support metacognitive strategies such as planning, observing and organizing, and improve the self-regulation skills of learners. For this purpose, Melanlioğlu (2016) states that rubrics have the purpose of teaching as well as assessment purposes.

In this study, as a part of process writing activities implemented to improve the writing skills of students attending B1 optional English preparatory classes in Muğla Sıtkı Koçman School of Foreign Languages, a rubric was developed (see Appendix A) to assess paragraph-level texts.

2. Method

This is a qualitative study in which the process steps of the rubric development were explained below. An analytical type of measurement tool was developed to assess paragraph-level writing texts, which were applied to improve writing skills of B1 optional preparatory students. Afterwards, validity, reliability and usefulness procedures were performed for the rubric.

2.1 Participants

In the development of the rubric; for the pilot study, opinions were taken from 60 university students and a lecturer with English language education from the optional B1 English Preparatory Program of Muğla Sıtkı Koçman School of Foreign Languages, and two lecturers from the English language education branch in order to test the compatibility between the evaluators, and within the scope of the reliability study in the stage of determining the dimensions and performance levels and criteria of the rubric, from three lecturers from the English language education branch, two experts in the field of educational sciences, and a lecturer from the Turkish branch for assessing in terms of grammar.

2.2 Development Process of the Rubric

At the stage of developing the rubric used to assess the achievement pre-test-post-test of process-based writing activities applied to improve paragraph-level writing skills of B1 level English preparatory students, process steps below were followed, which result from the synthesis of process steps used by Beyreli and Konuk (2016) in the grading key they developed to assess persuasive texts and process steps developed by Çıralı Sarıca and Koçak Usluel (2016) in the rubric they developed in order to be used in assessing digital storytelling in the educational context.

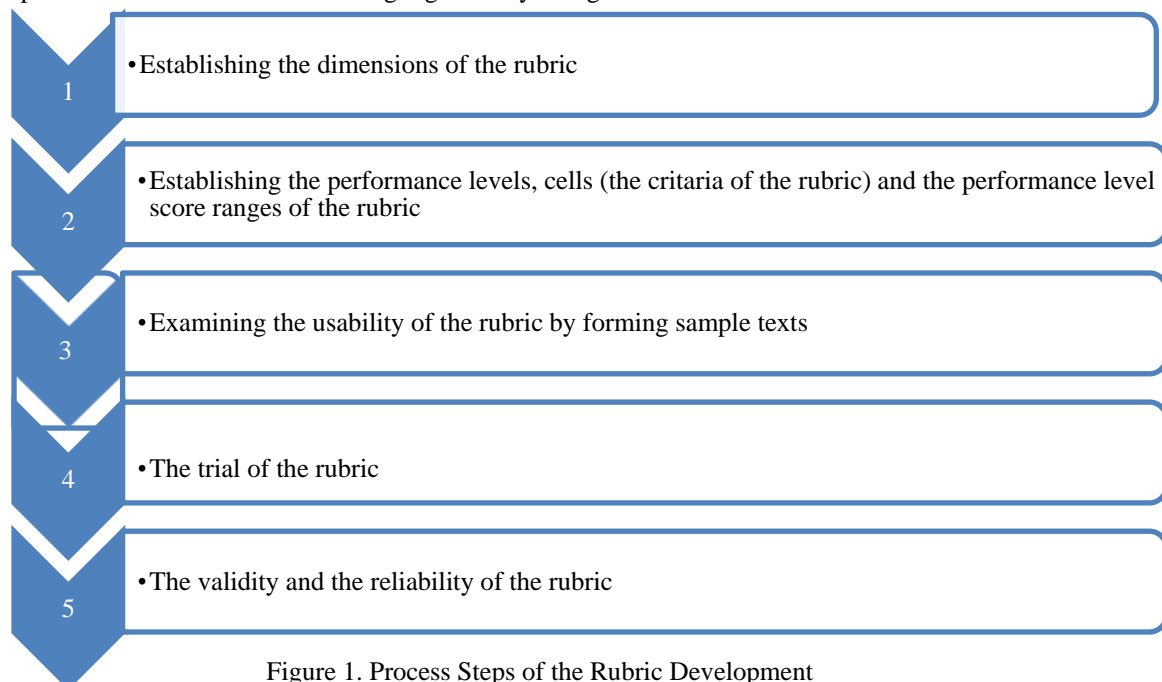


Figure 1. Process Steps of the Rubric Development

2.2.1 Establishing the Dimensions of the Rubric

As shown in Figure 1, the literature was examined regarding the types, dimensions, performance level and criteria of the rubric used in the assessment of writing skills in a foreign language, and the lines that are the dimensions of the rubric were determined. While determining the dimensions of the rubrics, initially, a criteria pool was formed by examining the rubrics used to measure writing skills in a foreign language and the criteria that were deemed appropriate were selected from this pool. The selected criteria were given short names related to the feature they measure, then the dimensions of the rubric were formed by grouping the criteria based on their similarities to each other. At this stage, the opinions of experts in English language education were sought.

2.2.2 Establishing the Performance Levels, Cells (The Criteria of the Rubric) and the Performance Level Score Ranges of the Rubric

In the decision-making process about how many performance levels the rubric will consist of in terms of scoring, the literature has been examined, various draft models have been reached, and after the expert opinions were taken, the 4-level model has been preferred. Cells, which are the criteria of the rubrics, were defined according to their performance level. During this procedure, it was regarded that the criteria would be distinctive. Repetitive readings were made to clarify the expressions and some expressions were changed. Simultaneously, the performance level score ranges of the rubric were determined.

2.2.3 Examining the Usability of the Rubric by Forming Sample Texts

During the usability stage of the rubric, a sample text was created by the researcher in accordance with the criteria defining the exact score that can be obtained according to the rubric, in order to crystallize the criteria that could seem abstract to the reader. By forming texts defining different performance levels on the same subject, the criteria were tested and the rubric was re-arranged during the process.

2.2.4 The Trial of the Rubric

During the rubric's trial phase, a pilot study was carried out with the students of B1 optional English preparatory class students who continued their education in Muğla Sıtkı Koçman School of Foreign Languages in 2016-2017 academic year, and within this study's scope, the texts written by the students were assessed in terms of usability with the rubric developed, and its deficiencies and defective aspects were determined and corrected.

2.2.5 The Validity and the Reliability of the Rubric

In the validity stage of the rubric, Moscal and Leyden (2000) recommend that the rubric should be addressed in terms of content, structure and criterion dimensions. Accordingly, in the content section, it is aimed to find evidence whether the assessment criteria include a foreign subject other than the subject being assessed or whether it includes all aspects of the content intended to be assessed; in the structure section, it is aimed to find evidence whether the assessment criteria can assess all parts of the structure or whether any assessment criteria are appropriate to the structure; and in the criterion dimension, it is tried to answer the questions about how the scoring criteria reflect their performance levels, what are the important components of the subject to be assessed with the rubric and how these components will be assessed or if there are any criteria remaining outside of the assessment. In this context, the draft rubric developed was presented to experts from the fields of foreign language, educational sciences and Turkish language, and reorganized in line with the feedback received.

Cohen's Kappa Coefficient, which is used to investigate the agreement between two raters, was used to calculate the inter-reader reliability in terms of the dimensions of the rubric. This test is a statistic used to measure the agreement between two readers in the evaluation of categorical items. Since the Kappa test also considers that the agreement between readers may be accidental, it is considered to be a stronger result than the agreement calculated as a percentage ratio between two observers. Kappa can have a value between -1 and +1. A Kappa value of +1 indicates a perfect agreement between the two observers, whereas -1 indicates that the disagreement between the two observers is excellent and 0 (zero) indicates that the agreement between the two readers is random. Landis and Koch (1977) suggest the data to be interpreted as; no agreement for a Kappa Score of 0 - 0.20, minimal agreement for 0.21 - 0.39, weak agreement for 0.40 - 0.59, moderate agreement for 0.60 - 0.79, strong agreement for 0.80 - 0.90, and almost perfect agreement for 0.90 and above.

3. Results

This part of the study includes findings related to the rubric developed to assess paragraph-level English writing texts, having written in order to assess the process-based writing activities conducted to improve the writing skills of B1 level English preparatory students. In this study, the analytical grading key type was chosen in order to minimize the subjectivity caused by the measurer and to give separate scores for different dimensions of the writing skill.

3.1 Establishment of Dimensions of the Rubric

During the process of creating dimensions of the rubric, a pool was formed as a result of literature review related to the assessment of writing skills. The establishment of the dimensions created in the process can be seen in the

Table 1. Establishment Stages of the Dimensions of the Rubric

The Beginning Stage of the Process	The Middle Stage of the Process	End of the Process
Content	Content	Content
Organization	Paragraph Structure	Paragraph Structure
Paragraph Structure	Grammar	Grammar
Sentence Structure	Vocabulary	Vocabulary
Grammar		Spelling and Punctuation
Vocabulary		
Spelling and Punctuation		

As shown in the table 1, within this criterion pool, dimensions that may be appropriate for the rubric originally developed were determined to be; content, organization, paragraph structure, sentence structure, grammar, vocabulary, spelling and punctuation. However, later in the process, at the stage of obtaining expert opinions and examining the texts written by the students within the scope of the pilot application, the organization dimension was removed because the dimensions of organization and paragraph structure overlapped with each other and it was difficult to distinguish the organization and the paragraph structure from each other in the tests performed on the sample texts. Similarly, since the sentence structure and grammar dimensions overlap, the grammar, which is a more comprehensive dimension, was preferred and the sentence structure was removed. In the texts examined within the context of the pilot study, it was found out that the students had almost no spelling and punctuation errors and that the scoring the spelling and punctuation dimension equally with other dimensions would not be right, so this dimension was removed. However, at the end of the research, it was decided that the spelling dimension should be absolutely included.

3.2 Establishing Criteria of Performance Level Criteria and Performance Level Score Ranges of the Rubric

As a result of the literature review, the fact that the rubrics found to assess writing skills have mostly 4-grade or 5-grade competency levels is attention grabbing. Since the definitions in the 5-grade competency level are very close to each other, it was seen that there was difficulty in distinguishing them and the 4-grade competency level was preferred within the scope of the research. Initially, the levels of competency were titled as higher than expected, meeting the expectation, not fully meeting the expectation and not meeting the expectation, but at the end of the process, the levels of competency was changed to "very good, good, needs improving and inadequate" because these were clearer expressions. In the stage of defining the criteria and determining the number of performance levels, the definitions of the highest and the lowest criteria were made first and the mid-level criteria were defined later. Various models have been developed in order to provide the distinction of the criteria and to eliminate the ambiguity between the criteria (See Appendix A). The formula [interval coefficient = (highest score - lowest score) / number of criteria] was used to calculate the performance level range coefficient to be used in the measurement process. The interval coefficient for the performance levels of the rubric was calculated to be 0.75. Accordingly, the performance level score ranges used to interpret the evaluation results are as follows: 1 - 1.75 = insufficient, 1.76 - 2.51 = needs improving, 2.52–3.27 = good, 3.28–4.00 = very good.

3.3 Creating Sample Text to Test the Usability of the Rubric

A sample paragraph text with criteria defining the full score in the rubric was created and presented to expert opinion. Afterwards, paragraphs with performance levels in 'sufficient' and 'needs improving' score ranges in the rubric were designed.

3.4 Trial of the Rubric

During the test of the rubric, an 8-week pilot study was conducted with 60 B1 optional English preparatory class students who continued their education in Muğla Sıtkı Koçman University School of Foreign Languages in the 2016-2017 academic year, and 60 paragraph-level texts written by the students within this study's scope were evaluated in terms of usability with the developed rubric, while trying to identify and correct its deficiencies and errors. In this process, it was decided that the spelling and punctuation dimension, which was determined in the initial rubric but removed during the development process, to be added back in, and that it should have the same level of competence with other dimensions in terms of preserving integrity and user friendliness.

3.5 Validity and Reliability of the Rubric

In the validity studies of the rubric, the opinions of the instructors who are experts in the field of English language education, the instructors who are experts in the field of educational sciences, and an instructor who is an expert in the field of Turkish language and literature were taken from the beginning of the process in the stages of forming its dimensions, criteria and performance levels, and the rubric has been revised many times.

Within the scope of the inter-rater reliability of the rubric, 60 independent texts at the level of paragraphs obtained from the pilot study applied to 60 students studying in the B1 optional English preparatory classes of Muğla Sıtkı Koçman School of Foreign Languages in 2016-2017 academic year were assessed and scored by two independent evaluators who are experts in the field of English language education, and Kappa coefficients were calculated. Inter-rater reliability results for the rubric were obtained. The results of the inter rater reliability analysis are given in Table 2.

Table 2. Weighted Kappa Scores of the Dimensions of the Rubric for the Inter Rater Reliability

	Context	Paragraph Structure	Grammar	Vocabulary	Spelling and Punctuation
K	.83	.81	.74	.74	.80
N	60	60	60	60	60
K= Weighted Kappa Score N= Number of Paragraphs Evaluated					

When Kappa scores regarding the dimensions of the rubric are examined in Table 2, it is seen that the agreement among the readers is significant, that the agreement is strong due to the fact that the Kappa score is .80 and above in the dimensions of content, paragraph structure and spelling, and that there's still agreement among the readers in the grammar and vocabulary dimensions but this agreement seems to be moderate due to the Kappa score being below .80. Moderate agreement can be interpreted as the effect of the number of categories of the variable on the Kappa score. The fewer the number of categories are, the higher the Kappa score calculated is (Viera & Garrett, 2005).

4. Discussion

In this study, the rubric which was developed in order to assess the paragraph-level texts written by B1 optional English preparatory students consists of 20 criteria and five dimensions which are "Content, Paragraph Structure, Grammar, Vocabulary and Spelling & Punctuation". Within the scope of the content study of the developed rubric, it was presented to the opinions of experts in the fields of foreign language, educational sciences and Turkish in various parts of the process and revised in line with their opinions. The weighted Kappa score between two raters for each dimension was calculated for the reliability of the rubric. Şencan (2005) states that when the Kappa score, which is the accord between raters, is higher than 60, it means that the reliability is fulfilled. In this context, when Kappa coefficients of the dimensions were examined, it was found that there was a moderate and strong agreement among the raters in terms of evaluation of all dimensions. Based on these results, it can be said that the rubric developed is valid and reliable.

5. Suggestions

Since the paragraph, which is one of the text types that can be produced by B1 level foreign language students in terms of writing skills, was preferred in order to assess the effect of the processes of writing activities within the study on writing success, the rubric prepared was developed for assessing the paragraph. However, there is a need to develop rubrics for different types of texts that students can write on all language levels, such as e-mail, story, digital storytelling or different types of articles.

References

- Atmaz, G. (2009). *Puanlama yönergesi (rubrik) kullanılması durumunda puanlayıcı güvenilirliğinin incelenmesi*. (Unpublished master's thesis) University of Mersin, Mersin Turkey.
- Bayram, F. (2006). *The role of writing portfolios in increasing learners' confidence in writing and promoting their attitudes towards writing*. (Unpublished master's thesis, University of Bilkent, Ankara, Turkey). Retrieved from <https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp>
- Beyreli, L., & Konuk, S. (2016). *Altıncı sınıf öğrencilerinin ikna edici yazma becerisinin geliştirilmesine yönelik bir araştırma*. Eğitim ve Bilim, 43(193). Retrieved from egitimvebilim.ted.org.tr/index.php/EB/article/view/7520
- Callison, D. (2000). Rubrics. *School Library Media Activities Monthly*, 17(2). 34-36,42.
- Çıralı, S. H., & Koçak, U. Y. (2016). Digital storytelling in the educational context: A rubric development study. *Eğitim Teknolojisi Kuram ve Uygulama*, 6(2), 65-83. <https://doi.org/10.17943/etku.12600>
- Chastain, K. (1988). *Developing second-language skills: Theory and Practice*. Orlando, Florida: HBJ.

- Gömlüksiz, M. N., & Elaldı, Ş. (2011). Foreign language teaching in the context of Constructivist approach. *Turkish Studies: International Periodical for the Languages, Literature and History of Turkish or Turkic*, 6(2), 443-454. <https://doi.org/10.7827/TurkishStudies.2418>
- Goodrich, H. (1997). Understanding rubrics: The dictionary may define “rubric,” but these models provide more clarity. *Educational Leadership*, 54(4), 14-17.
- Graham, S., & Sandmel, K. (2011). The process writing approach: A meta-analysis. *The Journal of Educational Research*, 104, 396-407. <https://doi.org/10.1080/00220671.2010.488703>
- Haladyna, T. M. (1997). *Writing test item to evaluate higher order thinking*. Massachusetts, MA: Allyn & Bacon.
- İnan, K. (2014). Yabancı dil olarak Türkçe öğrenen İranlıların yazılı anlatımlarının hata analizi bağlamında değerlendirilmesi. *Turkish Studies: International Periodical for The Languages, Literature and History of Turkish or Turkic*, 9, 619-649. <https://doi.org/10.7827/TurkishStudies.7319>
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130-144. <https://doi.org/10.1016/j.edurev.2007.05.002>
- Kan, A. (2007). Performans değerlendirme sürecine katkıları açısından yeni program anlayışı içerisinde kullanılabilir bir değerlendirme yaklaşımı: Rubrik puanlama yöntemleri. *Kuram ve Uygulamada Eğitim Bilimleri*, 7(1), 129-152.
- Kılıç, S. (2015). Kappa test. *Journal of Mood Disorders*, 5(3), 142.
- Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174. <https://doi.org/10.2307/2529310>
- Melanlıoğlu, D. (2016). A Metacognitive rubric for Turkish learners' listening skills. *Erzincan Üniversitesi Eğitim Fakültesi Dergisi*, 18(2), 1206-1229. <https://doi.org/10.17556/jef.20880>
- Mertler, C. A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research and Evaluation*, 7(25). Retrived from <http://PAREonline.net/getvn.asp?v=7&n=25>
- Moskal, B. M. (2000). Scoring rubrics: what, when and how? *Practical Assessment Research and Evaluation*, 7(3), 1-5. Retrived from <http://https://pareonline.net/getvn.asp?v=7&n=3&sa=U&ei=32>
- Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research and Evaluation*, 7(10), 71-81. Retrived from https://www.researchgate.net/profile/Jon_Leydens/publication/242637290_Scoring_Rubric_Development_Validity_and_Reliability/links/5857377208ae77ec37094497.pdf
- Nitko, A. J. (2001). *Educational assessment of student*. New Jersey, NJ: Merrill.
- Onozawa, C. (2010). A study of the process writing approach: a suggestion for an eclectic writing approach. *Research Note*, 153-163. Retrived from <http://www.kyoai.ac.jp/college/ronshuu/no-10/onozawa2.pdf>
- Parlak, B., & Doğan, N. (2014). Dereceli puanlama anahtarı ve puanlama anahtarından elde edilen puanların uyum düzeyleri. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 29(29-2). Retrived from <https://dergipark.org.tr/tr/pub/hunefd/issue/7788/101815>
- Popham, J. W. (1997). What's wrong-and what's right-with rubric. *Educational leadership*, 55(2), 12.
- Saddler, B., & Andrade, H. (2004). The writing rubric. *Educational Leadership*, 62(1), 48-52. Retrived from www.educationleader.com/subtopicintro/read/ASCD/ASCD_372_1.pdf
- Selvikavak, E. (2006). *Türkçenin yabancı dil olarak öğretiminde ileri düzeydeki öğrencilerin paragraf yazma becerisini geliştirme üzerine bir uygulama*. (Unpublished Master's Thesis, University of Ankara, Ankara, Turkey). Retrived from <https://tez.yok.gov.tr/UlusalTezMerkezi/tezSorguSonucYeni.jsp>
- Sezer, S. (2005). Öğrencinin akademik başarısının belirlenmesinde tamamlayıcı. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 18(18), 61-69. Retrived from <https://dergipark.org.tr/tr/pub/pauefd/issue/11125/133050>
- Stark, J. F. (1998). *Measurement and evaluation in education*. Ohio: Great & Great Publishers.
- Şencan, H. (2005). *Sosyal ve davranışsal ölçmelerde güvenirlik ve geçerlilik*. Ankara: Seçkin Yayıncılık.
- Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: The Kappa statistic. *Fam Med*, 37(5), 360-363. Retrived from www1.cs.columbia.edu/~julia/courses/CS6998/Interrater_agreement.Kappa_statistic.pdf
- Tekin, H. (2004). *Eğitimde ölçme ve değerlendirme*. Ankara: Yargı Yayınevi.

Appendix A. The Rubric Developed for Assessing Paragraph Level Written Texts

	PERFORMANS LEVELS			
	VERY GOOD 4	GOOD 3	MODERATE 2	POOR 1
1.Content	Content is fully relevant. It consists of sentences that are completely consistent and supports the main idea.	Content is generally relevant. With a few deficiencies, it consists of sentences that are generally consistent and supports the main idea.	Content is partially relevant. It doesn't fully support the main idea due to some inconsistent sentences.	Content is irrelevant. It consists of sentences that are inconsistent, and not supporting the main idea.
2. Paragraph Structure	Paragraph structure is fully relevant. There is a thesis statement, supporting sentences and a conclusion sentence.	Paragraph structure is generally relevant. There are a few deficiencies that doesn't affect the coherency of the paragraph.	Paragraph structure is partially relevant. There are some deficiencies that affect the coherency of the paragraph.	Paragraph structure is irrelevant. It is incoherent.
3.Grammar	Grammar is fully relevant. There are a lot of and complicated grammar structures.	Grammar is generally relevant. There are some different and complicated grammar structures.	Grammar is partially relevant. There are some repetitive simple grammar structures.	Grammar is generally irrelevant. There are a lot of grammar mistakes that make difficult to understand the paragraph.
4. Vocabulary	Vocabulary is fully relevant in terms of genre and meaning.	Vocabulary is generally relevant in terms of genre and meaning.	Vocabulary is partially relevant. There are some repetitive vocabulary mistakes that affect the meaning in the paragraph.	Vocabulary is generally irrelevant. There are a lot of genre and meaning mistakes that make difficult to understand the paragraph.
5. Spelling and Punctuation	Spelling and punctuation is fully relevant. (Punctuation, Capital letter, Spelling)	Spelling and punctuation is generally relevant.	Spelling and punctuation is partially relevant. There are some repetitive mistakes that affect the meaning in the paragraph.	Spelling and punctuation is generally irrelevant. There are a lot of repetitive mistakes that affect the understang of the paragraph.

* Do not assess paragraphs that consist completely irrelevant content or memorized texts.

* Assess paragraphs that consist insufficient content which is partially supporting the rubric.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the [Creative Commons Attribution license](#) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.