Journal of Ethnic and Cultural Studies
2022, Vol.9, No. 2, 121-133
http://dx.doi.org/10.29333/ejecs/1064

ISSN: 2149-1291

Assessment of Students' Argumentative Writing: A Rubric Development

Burhan Ozfidan[1]
Florida Gulf Coast University, USA

Connie Mitchell
Prince Sultan University, Saudi Arabia

Abstract: A rubric is used for assessing student work and performance. It is a tool that works in various ways to develop student learning and has great possibilities. The study presented aims to investigate the rubric development of second language learners' argumentative writing. The study's significance is to explore how well the rubric assesses students' achievement of the skills needed to develop argumentative essays. This study will add to the literature more data regarding rubrics' effectiveness in providing constructive feedback to students. This research describes the results of the current study in relation to rubric feedback from undergraduate students and the faculty who teach them from a private university in Saudi Arabia. The use of the rubric would be to support instruction and student performance. The researchers have proposed a methodology to design, develop, and implement a rubric as a scoring guide for argumentative essays based upon the achievement of learning outcomes for this genre. The rubric was developed to evaluate the following criterion: organization, integrating academic sources, thesis statement, finding evidence/lack of evidence, writing refutation paragraph, writing counterclaims, content, academic tone, mechanic, and vocabulary. The researchers statistically found significant interrater reliability and convergent validity coefficients. The results are considered to encourage the evaluation and development of such rubrics to be used across universities and colleges.

Keywords**:** argumentative writing, rubric development, learning outcomes, assessment.

Rubrics are a tool used to score student work and provide feedback at the same time. They consist of criteria and descriptors that describe the different levels of performance (Lee et al., 2020; Moskal, & Leydens, 2000). Furthermore, they provide students with a guide to what will be assessed, standards that need to be met, and information about where they are related to where they need to be. The genre of academic argumentative essay writing was selected as it is commonly used across higher education institutions (McCarthy et al., 2021; Reddy & Andrade, 2010). The need for a rubric that would provide supportive feedback on the students' achievement of this genre is an area that needed further development (Crawford et al., 2020; Curran et al., 2011; Trainor & Bal, 2014). They are an approach to scoring student work and have two primary methods analytic and holistic. A holistic rubric consists of a single integrated score assigned to a piece of writing.

[1] Corresponding Author: Assistant Professor of ESOL in the Department of Teacher Education at Florida Gulf Coast University. E-Mail: bozfidan@fgcu.edu

An analytic rubric consists of a set of criteria, with descriptors assigned to a piece of writing (Cabigao, 2021; Hyland, 2012; Mace & Pearl, 2021; Ene & Kosobucki, 2016).

Assessment in the field of writing especially in an environment where the students are studying in a foreign language requires preparation of well-written rubrics. The use of rubrics can allow students to receive the guided feedback they seek to improve their writing skills. Jeong (2015) highlighted studies on rubric use or its impacts on ratings are steadily increasing but are still at a primitive stage. This case study will add to the literature more data regarding rubrics' effectiveness in providing constructive feedback to students. This research will describe the results of the current study in relation to rubric feedback from undergraduate students and the faculty who teach them from a small private university in Saudi Arabia. Practitioners in the field of English Language need to develop skills and strategies on how to give feedback to their students about the achievement of the learning outcomes for a particular course. As we strive to build quality educational systems in higher education, the student's learning process needs to be at the forefront. Providing informative feedback will help students develop strategies to improve their learning and performance. Rubrics are a way to provide detailed feedback and help the students build knowledge of the areas they need improvement. Wolf and Stevens (2007) stated that "a rubric is a multi-purpose scoring guide for assessing student products and performances" (p. 1). This research will provide the results of a case study about how rubrics can provide constructive feedback and their impact on students' progress in achieving a particular course's learning outcomes.

A rubric is used for assessing student work and performance. It is a tool that works in various ways to develop student learning and has great possibilities (Stellmack et al., 2009; Tractenberg et al., 2010; Diab & Balaa, 2011). They can also be used to enhance teaching, play a part in the evaluation, and are an essential source of information for program development (Wolf & Stevens, 2007). The study discusses key terms of a quality rubric, presents a sample of a rubric for measuring a social science research study, and describes three basic steps in designing an efficient rubric.

Research Questions

The purpose of this particular study is to empirically develop an analytical rubric in order to assess students' achievement of learning outcomes in writing essay in the argumentative genre. The researchers used Halonen et al. 's (2003) study to empirically develop a rubric, which evaluates the success of certain learning outcomes using students' argumentative writing essays. The researchers were trying to discover how well the development of genre specific rubrics support student feedback and achievement of learning outcomes. The following were the research questions employed:

1. Did the rubric created for the study evaluate the success of the students' achievement of the learning outcomes for writing an argumentative essay?
2. Were the instructors using the rubric developed for this study able to score the essays without a significant difference using their method and the argumentative essay rubric developed by the researchers?
3. How well do genre specific rubrics support student feedback and achievement of learning outcomes?

Journal of Ethnic and Cultural Studies
2022, Vol.9, No. 2, 121-133
http://dx.doi.org/10.29333/ejecs/1064

ISSN: 2149-1291

Literature Review

Argumentative writing is a specific writing genre that highlights a position on an issue or topic and describes and supports this position with reliable pieces of evidence. This literature review highlighted the development process of a rubric, which directly reflects the study's findings.

Evaluate Learning Outcomes Using a Rubric

A rubric includes a descriptive list of the criteria that instructors assess to judge their students' study. Nitko (2001) affirmed that "rubrics are descriptive scoring schemes that are developed by teachers or other evaluators to guide the analysis of the products or process of students' efforts" (p. 132). Likewise, Chase (1999) highlights that rubrics as "scoring guides consisting of specific pre-established performance criteria, used in evaluating student work on performance assessment" (p. 32). Predominantly, a rubric for the written study comprises a list of definite aspects of writing performance, frequently subdivided under main categories, for example, organization, content, mechanics, language use, and vocabulary (Hack, 2015; Nimehchisalem et al., 2014). Scoring rubrics deliver a description of what is anticipated at each category or level with a view that students use the information to develop their upcoming performance. A rubric is firstly developed as an assessment tool only used by the instructors without notifying the students. In an effort to stress the powerful instructive elements of rubrics, Bangert-Drowns et al. (1991) stated, "rubrics are also teaching tools that support student learning and the development of sophisticated thinking skills" (p. 217). Additionally, Bangert-Drowns et al. (1991) indicated the strong link between teaching writing ability and the use of rubrics

> *it is usually used with a relatively complex assignment, such as a long-term project, an essay, or a research paper. Its purposes are to give students informative feedback about their works in progress and to give detailed evaluations of their final products*. (p. 218)

Holistic and analytic are types of rubrics that are identified in the literature consulted. Nitko (2001) highlighted that a holistic rubric forces a teacher to score the overall process or product as a whole without judging the component parts separately. In a holistic rubric, the focus of a score is on the overall quality, understanding, or proficiency of the particular content. An analytic rubric, according to Nitko (2001), "requires the teacher to score separate, individual parts of the product or performance first, then to add the individual scores to obtain a total score" (p. 54). An analytic rubric gives students specific feedback on their performance, and it is appropriate once there is a need to evaluate students' study in detail. An analytic rubric makes it potential to create a "profile" of specific student weaknesses and strengths. An instructor should decide whether the performance or product will be seen analytically or holistically before designing a specific rubric. Then, they should pilot the selected rubric as a form of instructional help.

Birky (2012) affirmed that a rubric provides great help to create informed learning/teaching settings and consciousness for the students. In a study, Wyngaard and Gehrke (1996) examined the relationship between improved writing skills and the use of criteria scales, using an analytic rubric including a clear description for each characteristic. These researchers investigated a rubric during a course and provided the students with the rubric to support them to evaluate their own studies. Using the same rubric, the researchers evaluated the students' studies at the end of the

implementation. They highlighted that the use of a rubric is an efficient way to develop student writing ability.

Jeong (2015) stated that "studies on rubric use or its impacts on ratings are steadily increasing but are still at a primitive stage" (para. 6). Based on this and a study conducted by Ozfidan and Mitchell (2020), the rubric development for argumentative essay writing is significant in providing feedback to the students and the teacher in order to support the students' development in the genre of argumentative writing assignments at university. Panadero and Jonsson (2013) have identified several aspects that support the effects of using rubrics. Furthermore, they stated that "rubrics may have the potential to influence students learning positively, but also that there are several different ways for the use of rubrics to mediate improved performance and self-regulation" (Panadero & Jonsson, 2013, p.135). They identified "a number of factors that may moderate the effects of using rubrics formatively, as well as factors that need further investigation" (p. 129). Jonsson and Svingby (2007) highlighted that rubric can increase performance assessments' reliability if they are more analytical, specific to the topic, and afforded the opportunity for rater training. They also pointed out that the rubrics appear to have the potential to become catalysts in "promoting learning" and/or "improving instruction" due to the explicitness of criteria and the facilitation of feedback and self-assessment.

Scoring Methods

Wind (2020) investigated whether raters using analytic rubrics for writing assessments were consistently marking their assessments. The purpose of the study was to inform rater training and develop any assessment of writing procedures. The study found that there was a lack of invariance for several of the participants in the study. Furthermore, according to a study conducted by Trace et al. (2016), scoring using rubrics and raters requires reliable methods to mitigate any differences among scores. They recommended using negotiation to reach an agreed-upon scoring system and agree upon the language used within the rubric. Their study utilized a mixed-methods approach to trace the implications of negotiation in the scoring decisions and how they were able to agree on the joint meanings of the rubric category criteria and descriptors. The results support the use of negotiation for raters to develop scoring inferences and shared meanings.

González et al. (2017) investigated the assessment of English as a Foreign Language (EFL) students' writing and the reliability of their scores. Their study revealed differences in terms of severity and leniency of the raters and the consistency of the scores even though the raters were from similar backgrounds. These factors need to be taken under consideration when setting out to develop rubrics and providing opportunities for the raters to have scoring sessions and discuss any issues with the language used and how it is interpreted in order to decrease the inconsistencies in scoring across raters.

Rubric Development

Zhao (2012) used a mixed-method approach in her study about the development and validity of using an analytic rubric that attempted to measure the voice strength for L2 learners in a piece of argumentative writing. Interestingly, her study revealed that the authorial voice in written discourse could be viewed through three criteria: "the presence and clarity of ideas in the content; the manner of the presentation of ideas; and the writer and reader presence" (Zhao, 2012, p. 208). Her study concentrated on the concept of voice in argumentative writing and how it can be assessed using a rubric. Dawson's (2017) emphasized that review of the literature on rubric development,

Journal of Ethnic and Cultural Studies
2022, Vol.9, No. 2, 121-133
http://dx.doi.org/10.29333/ejecs/1064

ISSN: 2149-1291

he came up with a framework with 14 design elements that need to be considered (see the Table 1 below).

Table 1
A Framework of a Rubric (Dawson, 2017, p. 352)

Element	Description
Specificity	the particular object of assessment
Secrecy	whom the rubric is shared with, and when it is shared
Exemplars	work samples provided to illustrate the quality
Scoring strategy	procedures used to arrive at marks and grades
Evaluative criteria	overall attributes required of the student
Quality levels	the number and type of levels of quality
Quality definitions	explanations of attributes of different levels of quality
Judgment complexity	the evaluative expertise required of users of the rubric
Users and uses	who makes use of the rubric, and to what end
Creators	the designers of the rubric
Quality processes	approaches to ensure the reliability and validity of the rubric
Accompanying feedback information	comments, annotation, or other notes on student performance
Presentation	how the information in the rubric is displayed
Explanation	instructions or other additional information provided to users."

These design elements are useful to consider when developing a rubric that will be used as a scoring guide on student work. Rubrics can make it easier to explain the results to students and others if warranted (Andrade, 2000).

Benefits of Using a Rubric

Chowdhury (2019) examined the process of creating a rubric that would support student feedback and as an instructional tool. The results indicated that rubrics do provide concrete feedback about the areas of strengths and improvement. The use of rubrics helped the students to evaluate their work. In another study. Turgut and Kayaoğlu (2015) conducted a quasi-experimental study that sought to examine the use of a rubric as an instructional tool to impact student learning. Interestingly, they found that using rubrics as an instructional tool was a way to enable students to become better writers. This improvement can be seen in their results since the experimental group outperformed the control group. Furthermore, they found that "when the teaching approach emphasized writing as a process rather than writing as a product after having internalized the rubric" (Turgut & Kayaoğlu, 2015, p. 56). Providing opportunities for students to be involved in the writing process and helping them to recognize the rubric as an instructional tool that is a guide for their writing process can benefit learners.

Research Method

Participants

The researchers used 145 students' (75 male and 70 female) argumentative writings. Consent forms were received from each student to use their papers in the current research study. This study was conducted in a private university in the Kingdom of Saudi Arabia. All participants' native language was Arabic. They all enrolled in ENG 103 (Research Writing Technique), and they were asked to write an argumentative essay as a final exam of this course.

Research Process

The students who enrolled in ENG 103 (Research Writing Technique) had to take a final exam at the end of the courses. The final exam of the course was to write an argumentative essay. The students were given a topic and asked to write an argumentative essay (approximately 650-800 words in length). They were asked to write five paragraphs. Firstly, they were asked to write an introduction to present background information on the given topic. Secondly, they were asked to write two different body paragraphs to represent two different ideas of the topic and support these ideas by detail and pieces of evidence. The students were given sources, and they were expected to use thesis sources to support their body paragraphs. Thirdly, the students were asked to write the third body paragraph, which was a refutation paragraph. The refutation paragraph had to identify the counterclaim and connect to the other side's main argument. Afterward, the students had to refute the counterargument and use evidence (using the sources) to support the refutation. Lastly, the students had to write a conclusion consisting of a summary of the final thought and restate the thesis statement. The rubric has ten criteria that assess the students' argumentative essay (Appendix A). Each criterion assesses a different perspective of writing an argumentative essay. Each criterion has a 5-point Likert scale: "Does not meet expectations (Mark 0), Below expectations (Mark 1), Developing expectations (Mark 2), Meets expectations (Mark 3), and Exceeds expectations (Mark 4)."

Procedures

The researchers initially developed a rubric to assess the students' writings. Halonen et al.'s (2003) study was used to experimentally develop a rubric, which evaluates the success of certain learning outcomes using students' argumentative writing essays. The rubric was reviewed by seven faculty members who were teaching ENG 103 (Research Writing Techniques). Each of the faculty members had experiences of teaching and assessing argumentative writings. The faculty members revised unclear terminology, vague statements, and the rubric's inappropriateness that might have been cause unfair assessment. According to the coders' feedback, the researchers also recruited two coders and revised the rubric two times. The coders coded all of the students' writings to assess interrater reliability. The raters were trained to use the rubric effectively before they used it. The researchers compared each essay's final ratings based on the rubric that each instructor used while evaluating students' written essays.

Journal of Ethnic and Cultural Studies
2022, Vol.9, No. 2, 121-133
http://dx.doi.org/10.29333/ejecs/1064

ISSN: 2149-1291

Interrater Reliability

To measure interrater reliability, the researchers used two raters to code written argumentative essays. The interrater reliability analyzed the two coders using Spearman's correlations, which is, according to McHugh (2012), a non-parametric test to measure the degree of association between two variables. Croux and Dehon (2010) stated "high inter-rater reliability values refer to a high degree of agreement between two examiners. Low inter-rater reliability values refer to a low degree of agreement between two examiners" (p. 501). Using a number of different statistics, inter-rater reliability for this study addressed the issue of consistency of the implementation of a rating system.

Convergent and Content Validity

Convergent validity, according to Carlson and Herdman (2012), "takes two measures that are supposed to be measuring the same construct and shows that they are related" (p. 18). Using Pearson correlations, the researchers analyzed convergent validity to measure the same constructs by comparing the rubric's criteria with the instructors' given scores of the course's argumentative writings. To grade the essays, the instructors utilized their own methods, which were not related to the developed rubric. All of the instructors provided the individual assignments that they have done previously.

Lawshe (1975) affirmed that content validity is essential in developing any new instrument. It provides evidence regarding an instrument's validity by measuring the degree to which the instrument assesses the targeted construct. For the study, content validity was conducted before the data collection. The researchers conducted content validity to enable the rubric to be used appropriate and meaningful inferences. Five experts in second language writing reviewed each item in the rubric for readability, clarity, and comprehensiveness. They cleared up some uncertain terms and made the rubric more understandable.

Results

Table 2 highlighted the researchers analyzed the interrater reliability between the two coders using Spearman's correlations. The interrater reliability average of the two coders was "rs (138) = .69, $p < .01$." The Vocabulary /Lexical diversity, "rs (138) = .92, $p < .01$", and the Content and Development criteria, "rs (138) = .88, $p < .01$", have the strongest correlations. The smallest correlation coefficient is "rs (138) = .34, $p < .11$", which is Writing Refutation Paragraph. The rest of the correlations between individual criteria ranged from .59 to .80 "see Cohen, 1988, for interpretations of correlation coefficients as related to reliability."

Using Pearson correlations, the researchers analyzed convergent validity. According to the rubric, the instructors' given scores of the essays from the courses compared to the essay's scores rated by the coders. While the instructors were grading the essays, they were used their own method, which was not related to the developed rubric. However, some of the instructors did not give the essay scores of some of the students. The results indicated that the rubric overall correlated with actual scores with a mean of "r (118) = .57, $p < .01$." Individual items varied; however, the thesis statement (preview to the argument) criteria, "r (118) = .51, $p < .01$," the integrating academic sources, "r (118) = .71, $p < .05$," the Mechanic (Grammar and Punctuation) criteria, "r (118) = .74, $p < .01$," and the Vocabulary /Lexical diversity criteria, "r (118) = .72, $p < .01$," all

yielded significant correlations. The researchers found that the rest of the rubric criteria are not significantly correlated (Table 2).

Table 2
Statistical Analysis

	M	Mdn	SD	IRR r_s	CV *rp*	α
Organization / Structure	4.1	4	1.1	.65**	.32	.81
Thesis statement (preview to the argument)	3.7	3	1.0	.61**	.51*	.84
Integrating academic sources	4.2	4	1.1	.70**	.71**	.79
Finding Evidence / Lack of evidence	3.6	3	1.2	.59**	.44	.75
Writing counterclaims	4.0	4	1.3	.80**	.29	.90
Writing Refutation Paragraph	2.3	2	1.0	.34	.45	.83
Academic tone	2.4	3	1.2	.78**	-.19	.86
Content & Development	2.6	3	1.0	.88**	.46	.73
Mechanic (Grammar and Punctuation)	3.5	4	1.0	.64**	.74**	.88
Vocabulary /Lexical diversity	3.2	4	1.1	.92**	.72**	.81

Note: "Mean, Medians, Standard Deviation, Inter-Rater Reliability Between the Two Coders, Convergent Validity, and Cronbach's alpha"
**Note*: $p < .05$
***Note*: $p < .01$
[1]*Note*: "Data does not include missing charts and represents 90% of the data. (138 out of 145) M"

The results of mean and SD scores of each item indicated that "Integrating academic sources" (M=4.2; SD=1.1), "Organization / Structure" (M=4.1; SD=1.1), and "Writing counterclaims" (M=4.0; SD=1.3) have the highest scores (see Table 2). The results also highlighted that “Writing Refutation Paragraph” (M=2.3; SD=1.0), “Academic tone” (M=2.4; SD=1.2), and “Content & Development” (M=2.6; SD=1.0) are have the lowest mean scores in the rubric. The rest of the item's scores were between 3.2 and 3.7. Table 2 also highlighted Cronbach's alpha value for each item in the rubric. Tavakol and Dennick (2011) indicated that the minimum Cronbach's alpha value should be .70 or above. Table 2 indicated that each item in the rubric represents a high Cronbach's alpha score (.73<items<.90). These findings highlighted that each item in the rubric is statistically reliable.

Discussion and Conclusion

A rubric includes a descriptive list of the criteria that instructors assess to judge their students' study. The researchers conducted this study to develop a rubric under the previous studies' guidelines. The researchers used Halonen et al. 's (2003) study to empirically develop a rubric, which evaluates the success of certain learning outcomes using students' argumentative writing essays. This developed rubric can be used to assess argumentative writings. Though there is a big struggle in clarifying the qualities that a strong manuscript's statements of hypotheses involve, the solid interrater reliability highlights that each criterion's meaning is sufficiently transmitted to the coders. Each criteria's low correlations could be attributed to the comparatively unclear nature of the criteria itself. The rest of the criteria in the rubric have solid correlations between the two coders. The researchers found that the criteria in the developed rubric are very similar to the actual

ISSN: 2149-1291

grade of the students' argumentative writing. The developed rubric might have difficulty capturing the instructors' interpretations of the students' writing and knowledge.

The process of creating a rubric that would support student feedback and as an instructional tool (Chowdhury, 2019). The results indicated that rubrics do provide concrete feedback about the areas of strengths and improvement. The use of rubrics helped the students to evaluate their work. In another study. Turgut and Kayaoğlu (2015) conducted a quasi-experimental study that sought to examine the use of a rubric as an instructional tool to impact student learning. Interestingly, they found that using rubrics as an instructional tool was a way to enable students to become better writers. This improvement can be seen in their results since the experimental group outperformed the control group. Furthermore, they found that "when the teaching approach emphasized writing as a process rather than writing as a product after having internalized the rubric" (Turgut & Kayaoğlu, 2015, p. 56). Providing opportunities for students to be involved in the writing process and helping them to recognize the rubric as an instructional tool that is a guide for their writing process can benefit learners.

This study statistically indicated significant interrater reliability and convergent validity coefficients. Using Pearson correlations, the researchers analyzed convergent validity. According to the rubric, the instructors' given scores of the essays from the courses compared to the essay's scores rated by the coders. While the instructors were grading the essays, they were used their own method, which was not related to the developed rubric. The findings are considered to encourage the evaluation and development of such rubrics to be used across universities and colleges. A rubric is used for assessing student work and performance. It is a tool that works in various ways to develop student learning and has great possibilities (Stellmack et al., 2009; Tractenberg et al., 2010). They can also be used to enhance teaching, play a part in the evaluation, and are an essential source of information for program development (Wolf & Stevens, 2007). The study discussed key terms of a quality rubric, presents a sample of a rubric for measuring a research study, and describes basic steps in designing an effective rubric.

Limitations and Future Studies

A limitation of the study is the raters coded each paper only one time. It would be better if the researchers could find more than two raters to code each paper more than two times. This study took a long time to collect and analyze the raw data since there were many participants. Therefore, the researchers had difficulties in finding appropriate raters to code each written argumentative essay. Since the researchers conducted various statistical analyses such as descriptive statistics (Mean, Median, and SD), interrater reliability, and convergent validity, all analyses should have been cross-checked by an expert statistician. Further researches should investigate other researchers' challenges and processes to develop their own assessment rubrics. Though we developed the rubric to assess argumentative writings, we will also develop another reliable and valid rubric to assess academic writing's general perspective for graduate writing-based courses.

Acknowledgment

This study was supported by Applied Linguistics Research Lab (ALLAB) at Prince Sultan University and Florida Gulf Coast University.

References

Andrade, H. G. (2000). Using rubrics to promote thinking and learning. *Educational Leadership, 57*(5), 13–19.

Bangert-Drowns, R., Kulik C., Kulik, J., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*, 213–238.

Birky, B. (2012). A good solution for assessment. *Strategies: A Journal for Physical and Sport Educators, 25*, 19–21.

Cabigao, J. (2021). Development and validation of a proposed model rubric in rating written outputs at the graduate school level. *Development, 5*(4), 122–131.

Carlson, K. D., & Herdman, A. O. (2012). Understanding the impact of convergent validity on research results. *Organizational Research Methods, 15*(1), 17–32.

Chase, C. (1999). *Contemporary assessment for educators*. Longman.

Chowdhury, F. (2019). Application of rubrics in the classroom: A vital tool for improvement in assessment, feedback and learning. *International Education Studies, 12*(1), 61–68. https://doi.org/10.5539/ies.v12n1p61

Crawford, A. R., Johnson, E. S., Zheng, Y. Z., & Moylan, L. A. (2020). Developing an understanding procedures observation rubric for mathematics intervention teachers. *School Science and Mathematics, 120*(3), 153–164.

Croux, C., & Dehon, C. (2010). Influence functions of the Spearman and Kendall correlation measures. *Statistical methods & Applications, 19*(4), 497–515.

Curran, V., Hollett, A., Casimiro, L. M., Mccarthy, P., Banfield, V., Hall, P., Lackie, K., Oandasan, I., Simmons, B, & Wagner, S. (2011). Development and validation of the interprofessional collaborator assessment rubric (ICAR). *Journal of Interprofessional Care, 25*(5), 339–344. https://doi.org/10.3109/13561820.2011.589542

Dawson, P. (2017). Assessment rubrics: Towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education, 42*(3), 347–360. https://doi.org/10.1080/02602938.2015.1111294

Diab, R. & Balaa, L. (2011). Developing Detailed Rubrics for Assessing Critique Writing: Impact on EFL University Students' Performance and Attitudes. *TESOL Journal, 2* (1), 52-72.

Ene, E. & Kosobucki, V. (2016). Rubrics and corrective feedback in ESL writing: A longitudinal case study of an L2 writer, *Assessing Writing, 30*, 3-20. https://doi.org/10.1016/j.asw.2016.06.003.

González, E. F., Trejo, N. P., & Roux, R. (2017). Assessing EFL university students' writing: A study of score reliability. *Revista Electrónica de Investigación Educativa, 19* (2), 91–103. https://doi.org/10.24320/redie.2017.19.2.928

Hack, C. (2015). Analytical rubrics in higher education: A repository of empirical data. *British Journal of Educational Technology, 46* (5), 924-927.

Halonen, J. S., Bosack, T., Clay, S., McCarthy, M., Dunn, D. S., Hill, G. W., IV, McEntarffer, R., Mehrotra, C., Nesmith, R., Weaver, K. A., & Whitlock, K. (2003). A rubric for learning, teaching, and assessing scientific inquiry in psychology. *Teaching of Psychology, 30* (3), 196–208. https://doi.org/10.1207/S15328023TOP3003_01

Hyland, K. (2012). *Second Language Writing*. Cambridge.

Jeong, H. (2015). Rubrics in the classrooms do teachers really follow them? *Language Testing in Asia, 5*, Article 6. https://doi.org/10.1186/s40468-015-0013-5

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity, and educational consequences. *Educational Research Review, 2*, 130–144.

ISSN: 2149-1291

Lawshe, C. H. (1975). A quantitative approach to content validity 1. *Personnel Psychology*, *28* (4), 563–575.

Lee, J. E., Recker, M., & Yuan, M. (2020). The validity and instructional value of a rubric for evaluating online course quality: An empirical study. *Online Learning*, *24* (1), 245–263.

Mace, M. K., & Pearl, D. (2021). Rubric development and validation for assessing comprehensive internationalization in higher education. *Journal of Studies in International Education*, *25*(1), 51–65.

McCarthy, K. S., Magliano, J. P., Snyder, J. O., Kenney, E. A., Newton, N. N., Perret, C. A., Knezevic, M., Allen, L. K., & McNamara, D. S. (2021). Quantified qualitative analysis: Rubric development and inter-rater reliability as iterative design. In *Proceedings of the 15th International Conference of the Learning Sciences-ICLS 2021*. International Society of the Learning Sciences.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*(3), 276–282. https://doi.org/10.11613/BM.2012.031

Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research, and Evaluation*, *7*(7),10-23. https://doi.org/10.7275/q7rm-gg74

Nimehchisalem, V., Chye, D. Y. S., & Jaswant Singh, S. K. A. (2014). A Self-Assessment Checklist for Undergraduate Students' Argumentative Writing. *Advances in Language and Literary Studies*, *5*(1), 65-80.

Nitko, A. J. (2001). *Educational assessment of students* (3rd ed). Merrill.

Ozfidan, B., & Mitchell, C. (2020). Detected difficulties in argumentative writing: The case of culturally and linguistically Saudi backgrounded students. *Journal of Ethnic and Cultural Studies*, *7*(2), 15–29.

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, *9*, 129–144.

Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, *35*(4), 435–448.

Stellmack, M. A., Konheim-Kalkstein, Y. L., Manor, J. E., Massey, A. R., & Schmitz, J. A. P., (2009). An assessment of reliability and validity of a rubric for grading APA-style introductions. *Teaching of Psychology*, *36*(2), 102–107.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

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

Tractenberg, R. E., Umans, J. G., & McCarter, R. J. (2010). A mastery rubric: Guiding curriculum design, admissions, and development of course objectives. *Assessment & Evaluation in Higher Education*, *35*(1), 15–32.

Trainor, A. A., & Bal, A. (2014). Development and preliminary analysis of a rubric for culturally responsive research. *The Journal of Special Education*, *47*(4), 203–216.

Turgut, F., & Kayaoğlu, M. N. (2015). Using rubrics as an instructional tool in EFL writing courses. *Journal of Language and Linguistic Studies*, *11*(1), 47–58.

Wind, S. A. (2020). Do raters use rating scale categories consistently across analytic rubric domains in writing assessment? *Assessing Writing*, *43*, 1–14. https://doi.org/10.1016/j.asw.2019.100416.

Wolf, K., & Stevens, E. (2007). The role of rubrics in advancing and assessing student learning. *The Journal of Effective Teaching*, *7*(1), 3–14.

Wyngaard, S., & Gehrke, R. (1996). Responding to audience: Using rubrics to teach and assess writing. *The English Journal*, *85*, 67–70.

Zhao, C. G. (2012). Measuring authorial voice strength in L2 argumentative writing: The development and validation of an analytic rubric. *Language Testing*, *30*(2), 201–230. https://doi.org/10.1177/0265532212456965

Notes on Contributors

Burhan Ozfidan is currently an Assistant Professor of ESOL in College of Education at Florida Gulf Coast University. He has completed his Ph.D. in ESL at Texas A&M University-College Station. He has extensive teaching and research experiences, and an educational background in applying multiculturalism, bilingualism, psycholinguistics, and language learning for educational purposes. ORCID ID: 0000-0002-2098-3409

Connie Mitchell has twenty-four years of teaching experience and is Education First's Teaching Excellence Award 2020 winner for Saudi Arabia, a Higher Education Academy Principal Fellow, and a TESOL-CAEP and NCAAA accreditation reviewer. At Prince Sultan University, she is the Vice Dean of the College of Humanities and Sciences and Director of the Teaching and Learning Center. ORCID ID*:* 0000-0001-9428-6035

Journal of Ethnic and Cultural Studies
2022, Vol.9, No. 2, 121-133
http://dx.doi.org/10.29333/ejecs/1064

ISSN: 2149-1291

APPENDIX A - Developed Rubric

Criteria	Does not meet expectations (Mark 1)	Below Expectations (Mark 2)	Developing Expectations (Mark 3)	Meets Expectations (Mark 4)	Exceeds Expectations (Mark 5)
Organization / Structure	There is no identifiable internal structure and readers have trouble following the writer's line of thought. Few, forced transitions in the essay or no transitions are present.	Arrangement of essay is unclear and illogical. The writing lacks a clear sense of direction. Ideas, details or events seem strung together in a loose or random fashion	Progression of ideas in essay is good. The writer sometimes lunges ahead too quickly or spends too much time on details that do not matter. Transitions appear sporadically, but not equally throughout the essay.	Overall, the paper is logically developed. Progression of ideas in essay makes sense and moves the reader easily through the text. Strong transitions exist throughout and add to the essay's coherence	Logical, compelling progression of ideas in essay; clear structure which enhances and showcases the central idea or theme and moves the reader through the text. Organization flows so smoothly the reader hardly thinks about it. Effective, mature, graceful transitions exist throughout the essay.
Thesis statement (preview to the argument)	No thesis statement	Thesis is completely unclear, not relevant, or missing entirely. No preview of argument. Paragraph consists largely of opinion, filler, or "contextualizing" material.	Thesis is largely unclear or is not directly relevant to assignment. Little identifiable preview, or preview is largely class material. Contains sentences that are neither preview nor thesis.	Thesis states a relevant position, but is somewhat vague or unclear. Preview is incomplete, not in correct order; some recitation of class material. Nothing other than thesis and preview.	Thesis clearly states a relevant position. Argument preview complete, in same order as body of essay, not simple recitation of class material. Contains nothing other than thesis and preview.
Integrating academic sources	Used little or no parenthetical documentation. APA in-text citations are not understandable. Paper dominantly uses informal language. Language is not appropriate to topic. Message is unclear.	Appropriate academic sources, but not a reliable. Source material is used, but integration may be awkward. All sources are accurately documented, but many are not in the desired format or lack credibility.	Majority of parenthetical documentation done incorrectly. Random APA citations throughout paper. Rarely documents sources. Some use of formal language recognized; informal language is frequently. Most language is appropriate to topic. Able to get vague idea of message.	Appropriate and reliable sources. Source material is used. All sources are accurately documented, but a few are not in the desired format. Some sources lack credibility.	Sources meet the guidelines for types of sources (APA in-text citations. Few errors noted in parenthetical documentation. Most research info is documented. Written in formal language (avoids slang completely). Language appropriate to topic. Words convey intended message.
Finding Evidence / Lack of evidence	Supporting evidence sentences lack development. Gives 1 or 2 weak claims that do not support that argument and/or are irrelevant or confusing. Does not include any paraphrasing from the packet texts.	Addresses few or none of the best arguments of the opposition. Includes some paraphrases from packet texts.	Addresses some of the best arguments of the oppositionIncludes a few good paraphrases from packet texts.	Supporting sentences clarify and explain the topic sentence. Includes some good paraphrases from packet texts.	Addresses the best arguments of the opposition. Includes good paraphrasing from packet texts. Paraphrasing supports student's ideas.
Writing counterclaims	Counterclaim is vague and includes little or no use of source material. Does not discuss the reasons against the argument or supporting claim.	Counterclaim not sufficiently clear and concise. Suggests there are reasons against the argument or supporting claim but does not discuss any specifically.	Counterclaim is stated sufficiently clearly and concisely that includes a few reasoned analyses and partial or uneven use of source material. Discusses weak reasons against the overall argument or supporting claim but leaves some reasons out and/or does not explain.	Counterclaim is stated quite clearly and concisely that includes some reasoned analysis and partial or uneven use of source material. Discusses reasons against the overall argument or supporting claim but leaves some reasons out and/or does not explain.	Counterclaim is stated very clearly and concisely that includes reasoned analysis and the use of source material. Discusses reasons against the overall argument or supporting claim and why it is invalid.
Writing Refutation Paragraph	Refutation is vague.	Some attempt at refutation is made but the points are not sufficiently developed.	A sufficient attempt at refutation is made but not all of the points are sufficiently developed.	The refutation is effective and the points are sufficiently developed.	The refutation is stated successfully countering it through evidence, whether it's evidence that conclusively disproves it by its findings or because it's more credible evidence.
Academic tone	No character of writer. Few connections between the narrative and the tone. It does not answer readers' questions. Invites little reader interest. Overuse of the passive voice.	Need to Improve academic writing. The tone is bland and does not fit the narrative. Needs to appeal to the readers more.	Character of writer varies. The tone strays from the narrative. Leaves some unanswered questions. Invites some interest. Moderate use of the passive voice.	Writing is in an appropriate academic tone. It is not distinctive and needs to engage the readers more.	Character of writer is consistently conveyed. The tone fits the narrative. Anticipates & answers readers' questions. Engages readers' interest and has an active voice.
Content & Development	Content is inconsistent and unclear.	Content is incomplete. Major points are not clear and/or persuasive appeals.	Content is not comprehensive and/or persuasive. Major points are addressed, but not well supported.	Content is comprehensive and/or persuasive. Major points are addressed, and somewhat supported.	Content is comprehensive and accurate. Major points are stated clearly and are well supported with good details. Content and purpose of the writing are clear.
Mechanic (Grammar and Punctuation)	Very weak spelling, punctuation, and capitalization. Incorrect grammar and sentence structure make it very difficult to understand the ideas.	Frequent errors in spelling, punctuation, and capitalization. Frequent errors in grammar.	Some errors in spelling, punctuation, and capitalization. Minor errors in grammar.	Mostly correct spelling, punctuation, and capitalization. A good command of grammar.	Only one or two errors in spelling, punctuation, and capitalization. A strong command of grammar.
Vocabulary /Lexical diversity	Significant use of incorrect and/or inappropriate vocabulary	Some wrong words.	Some wrong words.	Correct word choice and some usage of higher-level vocabulary.	Precise word choice and regular use of higher-level vocabulary.