Published at https://ijate.net/   https://dergipark.org.tr/en/pub/ijate   *Research Article*

Gathering evidence on e-rubrics: Perspectives and many facet Rasch analysis of rating behavior

Inan Deniz Erguvan [ID][1,*],   **Beyza Aksu Dunya** [ID][2,3]

[1]Gulf University for Science and Technology, Block 5, Building 1, Mubarak Al-Abdullah Area, West Mishref, Kuwait
[2]Bartin University, Bartin, Turkey
[3]University of Illinois at Chicago, Chicago, IL, USA

Abstract: This study examined the faculty perspectives towards the use of electronic rubrics and their rating behavior in a freshman composition course. A mixed-methods approach has been employed for data collection and analysis. The data for faculty perspectives were collected from nine instructors through semi-structured interviews and for their behavior, six instructors teaching the same course in Fall 2019, shared their students' essay scores with the researchers. Many facet Rasch model (MFRM) was employed for quantitative data analysis. According to the findings of the quantitative data, the instructors differed in their degree of leniency and severity, one instructor being more lenient and one being more severe than the others. Another interesting finding was one instructor turned out to be an inconsistent user of the e-rubric. The findings of the qualitative data showed that writing faculty think e-rubrics come with great advantages such as facilitating scoring, ensuring standardization, and reducing student complaints and grade appeals. However, they view the impact of e-rubrics on student writing with cautious optimism. The findings of the qualitative and quantitative strands are overlapping, and the responses elicited from the participants seem to shed some light on the rating behavior of the writing faculty.

1. INTRODUCTION

The word "rubric" implies an assessment tool that describes levels of performance on a particular task and is used to assess outcomes in a variety of performance-based contexts (Hafner & Hafner, 2003). Rubrics, by this definition differ from rating scales, which have criteria but no performance level descriptions, although these may be confused with "rubrics" (Brookhart 2013). Rubrics, checklists, and rating scales all have criteria but what distinguishes them is the scale. Other than rubrics, none of the other scales offer students a description of the quality of their performance they can easily use to envision their next steps in learning (Brookhart, 2018).

*CONTACT: Inan Deniz ERGUVAN ✉ erguvan.d@gust.edu.kw ▣ Gulf University for Science and Technology, Block 5, Building 1, Mubarak Al-Abdullah Area, West Mishref, Kuwait

There are many benefits of using rubrics; providing consistency of scoring across students, assignments, and different raters is one of the major benefits of using rubrics. Designing and using rubrics to grade assignments or tests can indeed reduce inconsistencies and make grading of the written work more objective. Subjectivity is a big concern in assessing writing, and the use of rubrics can help remove bias from grading (Allen & Tanner, 2006). Rubrics indeed provide an opportunity for reliable scoring, rather than a subjective scoring simply based on the rater's personal idiosyncrasies (Carr, 2000).

Sharing the rubric with students can have the added benefit of enhancing learning by allowing for feedback and self-assessment (Jonsson & Svingby, 2007; Reddy & Andrade, 2010). The rubric tells both teachers and students what fundamental skills teachers look for while they are assessing student performance (Arter & McTighe, 2001) because they incorporate criteria to rate the essential dimensions of performance, as well as standards of achieving those criteria (Jonsson & Svingby, 2007).

Another advantage of using rubrics is facilitating self- and peer-evaluation, both of which could be valuable avenues for providing meaningful feedback. With the development of a simple rubric, students can peer-review each other's work, thus see other examples other than their own and that of the instructor. Perhaps most importantly, sharing rubrics with students can support them in identifying where their thinking has gone wrong and promote learning (Jonsson & Svingby, 2007; Reddy & Andrade, 2010). As Broad (2000) suggested, when rubrics are utilized accurately, learners not only get feedback from the instructor, but they also receive training in self-assessment. Self-evaluation of one's work using the instructor's rubric can build meta-cognitive skills in making self-corrections. Such evaluations may provide meaningful feedback that could further enhance the learning process (Sadler & Good, 2006; Freeman & Parks, 2010). However, Brookhart and Chen (2015) underline the fact that rubrics that include descriptions of quality on criteria that reflect learning goals, rather than rubrics that focus on the requirements for an assignment and not indications of learning function as the goals toward which students can monitor their progress,

While the attitude towards using rubrics is prevalently positive, there are some negative perspectives, too. Critics complain that rubrics are rigid and even when they are modified to allow for more commentary on student strengths and weaknesses; they do a disservice to students' ability to learn. Critics also add that rubrics result in standardized measurement of standardized writing, which is not the purpose of writing instruction (Nordrum et al. 2013; Torrance, 2007).

Andrade (2000) brings up in a study that rubrics are not necessarily self-explanatory and not all students are acquainted with rubrics. Therefore, teachers must not assume that the criteria in the rubric are all clear to students. Andrade (2005) also alerts that rubrics must pass a test of quality, demonstrating that if another instructor utilizes the same rubric to review the same paper; their results should only have insignificant differences.

Some critics are concerned about the impact of using rubrics on creativity of students. Kohn (2006) argues that rubrics result in student writing with less depth of thought, therefore rubrics should not drive instruction, nor should they be shared with students. Kohn also says an excessive amount of consideration regarding the nature of work causes the students to lose enthusiasm for whatever they are doing. Another critic of rubrics, Wilson (2007), suggests that over-reliance on rubrics may result in learners stopping writing for a live audience, and beginning to write for a rubric. Wilson (2007) argues that the rubrics provide students with non-specific input and do not have much relevance with what they need to say and adds that writing offers its own set of criteria and that each piece should be examined individually.

To mitigate most of these issues, some studies emphasize the importance of involving students in developing rubrics and reduce the number of criteria incorporated so that they become easier to comprehend and apply as a learning tool (García-Ros, 2011).

1.1. E-rubrics

With the rise of the digital age, Information and Communications Technologies (ICTs) have started offering new roles and resources to teachers as well as students to improve teaching and learning processes. As technology started being widely used in assessment techniques, standard rubrics have slowly been replaced by their digital companions, called electronic rubrics (e-rubrics) (Raddawi & Bilikozen, 2018). Simply put, electronic rubrics or e-rubrics are rubrics that are presented online.

E-rubrics carry out the same functionality as paper or print-out rubrics, but there is an added value of e-rubrics. According to Steffens (2014), students are increasingly working in technology enhanced learning environments. From a technical point of view, it is relatively easy to integrate e-rubrics in technology enhanced learning environments. Using e-rubrics has the advantage that feedback can be given much more quickly than in traditional learning environments with paper-and-pencil. Just like any computer-assisted system, e-rubrics make grading and assessing much simpler for instructors as they reduce the time required to grade assignments. E-rubrics facilitate more immediacy in the teacher - student communication, and frequent and quicker feedback may help students to better self-regulate their learning than in traditional learning environments (Rivasy et al., 2014), and to be active participants in the learning process.

Kirwin and DiVall (2015) express that e-rubrics offer different advantages to various groups such as students, course instructors and administrators. Students can use the feedback and comments within a rubric as well as the scores on particular items to see their strengths and weaknesses. Course instructors can use e-rubrics to see the dimensions of performance and aggregate across multiple assignments to examine learning outcome data. Program administrators can benefit from e-rubrics by aggregating student performance as an indicator of the group's competency in a particular area. With the help of learning management systems (LMSs) that are capable of aligning course outcomes with particular dimensions in e-rubrics, it will be quite practical to evaluate learning outcomes for a large number of students.

Another benefit of e-rubrics was put forward by Martinez et al. (2016). In this case study, the course professor and students generated a collaborative methodology to build a rubric with the support of educational technologies. As a result of the collaborative effort, the students and professor agreed on the criteria for assessment of student presentations. This effort eventually showed that the evaluations given by students and the course professor got closer thanks to the increased e-rubric use. E-rubrics have the advantage of facilitating collaborative rubric generation among course professors and students, which could reverse the drawback of rubrics as argued by some critics that rubrics are not always clear to students.

However, one must acknowledge the fact that writing assessment is a complex and error-prone cognitive process. Therefore, attention should also be turned to raters themselves because in the end what is central to writing performance assessment is the rater behavior. Researchers have long recognized that rater judgments have an element of subjectivity. It is inevitable that the act of rating involves rater errors or rater biases (Myford & Wolfe, 2003), and although raters are trained to use and interpret rating scales in similar ways, rater effects also need to be studied. Rater behavior must be taken into consideration in order to assess the construct in question. Among many potential rater errors, four major categories of rater errors have been given emphasis: (a) severity or leniency, (b) inconsistency (c) halo, and (d) restriction of range (Myford & Wolfe, 2003; Saal et al., 1980).

In this study, we focused on two common rater errors: Severity/leniency and inconsistency. The former is defined as the tendency of a rater to assign higher or lower ratings on average than those assigned by other raters, and it is commonly considered to be the most pervasive and detrimental rater effect in performance assessments (Dobria, 2011). Various factors contribute to a rater's severity or leniency including professional experience, and in some circumstances, the most experienced or senior rater may also be the most severe (Eckes, 2011).

Rater inconsistency is a rater's tendency to apply the rating scale inconsistently compared to the way other raters apply the same scale. The presence of rater inconsistency indicates the rater's lack of understanding of rating criteria, making the interpretation of ratings less meaningful. A rater who rates inconsistently increases the randomness in scores by assigning high ratings to those who deserve low ratings and low ratings to those who deserve high ratings. This error reduces the ability of the scores to reliably differentiate between competent and incompetent students (Iramaneerat & Yudkowsky, 2006).

Thus, the aim of this paper is double fold: to analyze the perspective of writing faculty towards using e-rubrics through interviews and to examine their rating behavior via Many Facet Rasch Model (MFRM) in a first-year composition course in a university in Kuwait. MFRM is a member of Rasch Measurement Models that is suitable to simultaneously analyze multiple facets potentially having an impact on scores (Eckes, 2011). It is an extension of the basic Rasch Model for analyzing dichotomous data and used in assessments that involve human judgment. It allows researchers to investigate potential sources of error that cause construct irrelevant variance into the ratings. The advantages of MFRM also includes that each facet's unique contribution to the scores can be partitioned out and investigated independently of other facets in the assessment (Myford & Wolfe, 2003).

The use of LMS integrated online rubrics in the institution dates to 2015, when the practice of using a common analytic rubric was adopted. A common rubric is used in all writing courses which is developed and revised by the course coordinator every year. Writing faculty upload this common rubric in Turnitin in order to check the plagiarism similarity index of the essays and also to mark and give feedback to their students' essays. However, standardization workshops conducted during the academic year indicated some discrepancy among instructors, thus, the researchers decided to conduct a study during the Fall semester of 2019-2020 to analyze whether the severity / leniency is a real problem in the department. The qualitative component was added to analyze the instructors' perspective towards the use of the e-rubric.

To this end, this paper tried to find answers to the following research questions:

1. What are some benefits of using e-rubrics in an academic writing class?
2. What are some limitations of using e-rubrics in an academic writing class?
3. How do e-rubrics compare to rubrics in assessing writing?
4. Does using e-rubrics affect students' writing performance positively?
5. Do the instructors differ in terms of their level of severity while rating the student essays with the standard e-rubric? If yes, which rater is more severe / lenient than others?
6. How consistently are the instructors able to distinguish among the students in terms of their levels of proficiency?

2. METHOD

This study used a mixed method research design for data collection and analysis. In the mixed method research design, both qualitative and quantitative methods are used and "mixed" for collecting and analyzing data in a single study (Creswell, 2012). In mixed method research design, the two forms of data are mixed concurrently or sequentially by giving priority to one or both forms of data (Creswell & Clark, 2011). Using multiple methods helps to provide a

more comprehensive framework of the phenomenon under investigation by enabling rich and informative data and also to validate and triangulate the data by analyzing the same issue through both quantitative and qualitative methods (Silverman, 2000).

For the qualitative study, the researchers conducted semi-structured interviews with the writing faculty. The interview questions were developed to evaluate the raters' perspective towards e-rubrics. The researchers adapted Raddawi and Bilikozen's (2018) interview questions to evaluate ELT professors' perspectives on the use of rubrics in an academic writing class in a university in the United Arab Emirates.

2.1. Research Population

The data used for the qualitative strand of this study came from the interviews conducted with the writing faculty working in the university in January to March 2020. Nine instructors teaching various writing courses in the Spring 2020 academic year were interviewed for their perceptions of using e-rubrics in assessing their students' essays. Four of the faculty members were native speakers of English, whereas five of them were non-native. Out of nine, two were female, and seven were male.

The data used for the quantitative strand of the study came from six instructors teaching a particular writing course in the Fall 2019, in the same university in Kuwait. Out of six instructors, four of them were male, two were female; three were native speakers, three were nonnative. The total number of students taking this particular writing course is 442, and the number of essays that were rated by the six instructors is 424.

2.2. Data Collection

The qualitative strand of the study involved interviewing the writing faculty members in the department. According to Brown (2001), interviews have a high return rate and fewer incomplete answers. They also allow researchers to ask for clarification in a participant's response to a given question (Mertler, 2009) As a result, interviews offer an advantage over surveys as researchers can get more details on vague answers.

To protect the anonymity and confidentiality of the participants, each was given a code. As for the reliability and validity of the interviews, the researcher conducted some pilot interviews with some faculty who are not in the sample to check the understandability of the questions. Due to the interactive nature of the interview and the various biases and limits that may impact human decision-making, during the interviews the interviewer did not deviate from the interview questions and kept a neutral body language with all interviewees. After the interview, the recorded voice file and the written interview text were sent to each interviewee to obtain their approval to avoid any misunderstandings.

For the quantitative strand, six instructors scored the final draft of their students' research-based essays using a common analytic e-rubric. The common e-rubric is uploaded on Turnitin and attached to the writing task; therefore, scoring takes place electronically (See Appendix 1 for the e-rubric). Student essays were rated by the instructors during the first week of December 2019. No special training or a norming session was provided prior to or during the rating process. The instructors then shared their students' scores with the researcher. Also, for anchoring purposes, the researcher randomly selected two essays from the pool of essays to be rated by the six writing faculty members. The common frame of reference made it possible to compare all students and all instructors on the same scale.

2.3. Data Analysis

The data analysis process that could be utilized in qualitative research can be broken down in three steps: preparing and organizing the data for analysis, reducing the data into themes through coding and condensing the codes, and finally representing the data in figures, tables,

or discussion (Creswell, 2007). In this study all interviews were audio-recorded and fully transcribed in order to prepare them for the analysis. Following Radnor's (2002) approach to analyzing semi-structured interviews in interpretive research, the data was further prepared for analysis by reading the transcribed interviews several times and noting down the topics emerging from the data. During the data analysis process, we read the transcripts carefully to draw out any implicit topics that we may have missed. We made a list of the topics and gave a code to each topic. Afterwards, we pulled out the categories within each topic and listed these categories under each topic as subheadings. We also counted the frequency of these categories in interview texts to indicate which categories are more commonly expressed by the interviewees. The percentage of agreement between the coders which represents the share of common number of codes with respect to the total number of codes was calculated as a measure of consistency for coding. As a rule of thumb, 80% agreement between coders is sufficient for ensuring intercoder reliability (Miles & Huberman, 1994). The percentage of agreement between coders was found 92% for this study. This value was well above the suggested value of 80%. The next step was reading the transcripts for content, which meant going through the text one more time to highlight and code the main quotes that go under each category. These quotes were used to illustrate the participants' voices and viewpoints more clearly in the discussion of the findings.

For the quantitative strand, we employed the MFRM (Linacre, 1989) to analyze the rater behavior by using FACETS program (Linacre, 2020). A facet is an aspect of any assessment situation that may have an influence on the outcome. A facet can be raters, performance tasks, or examinee-related characteristics such as race, gender, etc. (Myford & Dobria, 2012). In FACETS output, a column titled "measure" displays each instructor's severity measures in log-odd units. These measures are estimates of each rater's true location on the severity dimension (Eckes, 2011). MFRM can separate out each facet's unique contribution to the assessment setting and examine it independently of other facets to determine to what extent each facet is functioning. The advantage of MFRM with respect to classical approach while examining rating data is that MFRM allows an in-depth analysis of patterns in ratings even when a different set of examinees are concerned. In the classical approach, interrater reliability is reported while analyzing rating data. Interrater reliability is an informative statistic, yet it is limited in detecting possible rater effects such as severity/leniency. MFRM provides a valid account of potential irrelevant variance sources in ratings such as severity/leniency or bias. There are multiple indicators for detecting rater effects under MFRM framework which includes outfit and infit mean-square indices (Myford & Wolfe, 2003). Infit and outfit indices are used to assess randomness in the scores assigned by raters. These values are averages of squared standardized residuals and have an expected value of 1.00. Specifically, mean-square outfit which is more sensitive to outliers in the data is the non-weighted mean of the squared standardized residuals while infit is the information-weighted mean of the standardized residuals (Wolfe, 2009). In both statistics, values greater than 2.00 are accepted as indication of severe misfit that distort the measurement (Linacre, 2009). FACETS program yields individual fit values for each rater to assess rater misfit. In this part of the analysis, raters who had fit indices greater than 2.00 were flagged for further review.

3. RESULTS

3.1. Qualitative Study

Nine faculty members who were interviewed come from various backgrounds and nationalities, as displayed in Table 1. The table displays that although the writing faculty in the university are quite varied in their nationalities, they have similar backgrounds in education and experience in teaching. Their experience in teaching is also reflected in their experience with using rubrics, with eight of them having more than eight years of experience using rubrics in

assessing student essays. Understandably, the experience with e-rubrics is at an average of three to four years, simply because e-rubrics are new tools in assessing writing and they have been recently adopted in writing courses with the accelerated integration of internet technologies in education.

Table 1. *Demographics of interviewed faculty members.*

Interviewee	Nationality	Qualification	Teaching Experience	Experience with rubrics	Experience with e-rubrics
P1	American	MA in English Literature	35 years	20 years	14 years
P2	Kuwaiti	PhD in Linguistics	4 years	8 months	8 months
P3	Kuwaiti	MA in English	10 years	10 years	3 years
P4	Indian	MA in English Literature	25 years	20 years	15 years
P5	British	MA in ELT	25 years	10 years	4 years
P6	Egyptian	PhD in Composition & Rhetorics	10 years	10 years	4 years
P7	Bosnian	PhD in TESOL	21 years	10 years	6 months
P8	American	MA in TESOL	8 years	8 years	3 years
P9	New Zealander	PhD in TESOL	24 years	20 years	3 years

When asked what they used for grading and giving feedback before they started using rubrics, two faculty members said they do not remember any time when they did not use rubrics for grading. The remaining three faculty members said they graded holistically with plenty of qualitative feedback and four of them said they used a previously agreed upon checklist, scheme and some standards based on learning objectives of the course. Basically, even the instructors who used to grade holistically always referred to some standards, attended some standardization sessions or had a checklist to refer to, which means most of the writing faculty had exposure to a scorecard for grading before they fully adopted rubrics in their classes.

3.1.1. *Research question 1*

When the faculty members were asked what they see as the greatest benefit of using e-rubrics in assessing student essays, they gave a variety of responses which could be summarized in Table 2.

Table 2. *Perceived benefits of e-rubrics.*

Perceived benefit	Frequency
Making grading objective and transparent	6
Reducing student complaints / grade appeals	4
Facilitating grading	4
Helping standardization among raters	3

Table 2 shows that the most frequently mentioned benefit of using rubric is making grading transparent to students as they clearly show how their essays are graded and where their weaknesses and strengths are. Another benefit mentioned quite frequently, by four instructors, is "rubrics reduce complaints". In fact, these two perceived benefits are quite closely connected to each other, because instructors feel less defensive when they can explain clearly where the students' scores come from and they are able to "justify" the scores.

P6. I would say the biggest advantage is that they break down the grade in a very quantifiable easy to explain manner so the student would see exactly how they received that grade. They

would have a clear idea about the criteria on which they were graded. They would know exactly what their strengths and weaknesses on each criterion are or were.

Some instructors see an added benefit to making grading transparent in reducing the complaints and negotiations:

P5. I make strenuous efforts to get the students to review the e-rubrics… as whole class activities before major essays so they are constantly aware of the requirements. That being the case, no student can make any excuses regarding being under graded or being sort of penalized in some way because they are all aware of what the requirements are, and they know that if they don't meet these requirements they have no complaints. Students have a clear template of how they got their grades. It clarifies everything for everybody. It reduces conflicts. It reduces grade appeals because, theoretically, there is nothing subjective about a rubric. This creates an objective way of grading and it roughly solves all issues and answers all questions.

It seems student complaints due to scores are a problem in the institution and rubrics are useful tools to offer relief to instructors in this regard.

P9. I show the whole class how I am grading... I put the rubrics here and I point out this is why he is getting 80 here for the grammar, for example, because he has made these grammar mistakes or the sentences are not complex enough, so I have chosen 80 on this rubric but they don't really read the actual all the dots on the rubric. They just want the general grade. When I grade three to five sentences that had grammar errors or such and such, they don't care about that, but they get the general idea about the grading.

Rubrics helping standardization among raters has also been mentioned by 3 instructors. For the interviewee 7, quoted below, that was second major advantage of rubrics, after making grading transparent for students:

P7. The biggest advantage is that … It has actually two advantages. First, they are used to the system. Once you do this once, they get used to how they're being graded. Second, it is standardized across the university so, you know, not this professor is better than this professor because they do not grade in the same way.

Facilitation of grading was mentioned by three instructors, and they mentioned the benefit of justification along with the ease of grading.

P3. I am thinking about if you have large number of students in class who are trying to take the same assessment you have 60 to 90 papers to get through, it does facilitate the grading process and also facilitates the feedback process because going back to every single paper I would forget how and why I graded this paper and gave it that specific grade but going back and looking at the rubric itself and you kind of have that sliding option on Turnitin, for example, it does make easy for me to provide feedback and justify the grade as well as, when it comes to the students, it helps the students sync in, I guess.

3.1.2. *Research question 2*

When the faculty members were asked what they see as limitations of using e-rubrics in assessing student writing, they gave a variety of responses categorized under three themes, which is shown in Table 3. As the table displays, the perceived limitations are mainly related to scoring problems and the problems the students have while using the e-rubric, as well as time and effort the construction of a successful rubric requires. Besides, two participants expressed they cannot think of any limitations related to e-rubrics, and this was also displayed in the table as a separate theme.

Table 3. *Perceived limitations of e-rubrics.*

Perceived limitation	Frequency
Scoring issues (ambiguity, rigidity)	4
Student issues (accessing / reading and understanding the e-rubric)	4
Construction of the rubric	2
No limitation	2

The perceived limitation "scoring issues" is composed of a couple issues as defined by instructors. Some have pointed out the limitations of rubrics in differentiating certain categories from one another and some have brought up the lack of flexibility when it comes to issues such as giving half points in the rigid scoring scheme of the rubric. However, these points are not necessarily related to the nature of e-rubrics in particular, but rubrics in general.

P6. *I would say the disadvantage would be that it would be very difficult to design a rubric that would cover all the possible mistakes and any potential drawbacks on any given criteria. There is no category or there is no description on the rubric so I would say that the challenge is that there is always something missing in the rubric because no matter how detailed and descriptive you are, you are always not going to be able to cover everything on a one-page rubric ...so the categories are not always clear cut in reality as they are presented on a rubric ... the criteria may fall between two categories ...*

P3. *Certain rubrics especially ones that are weighted in a specific way it would be hard to differentiate, for example, between grammar mechanics from content so if the content is preferred but the grammar is being marked down then grammar shouldn't be as important but if the grammar hinders the actual sentence and the structure and makes it completely incomprehensible then I, for example, run the risk of marking them either too harshly or too leniently.*

Three instructors mentioned the problems their students go through while accessing the e-rubrics, and the fact that they never even try to access the e-rubrics. This instructor quoted below describes the limited accessibility problem of e-rubrics for some students:

P7. *... I would say the students' ability to access the feedback. This is not related to our courses. This is rather related to their IT skills.... If it were for me, I would ask them to be trained by IT before they ever access our classes... Their access to e-rubrics is limited because of their lack of knowledge of how to navigate this technological thing, I mean the use of computers. Many of them, I discovered at the end of the year, could not access their feedback once and I had highlighted word by word almost then...in addition to the rubric. If it were a paper rubric, they would have it under their noses.*

Another student-related problem is students not reading the rubrics. This could be related to their lack of IT (information technology) skills or lack of interest in reading and or improving themselves in the course.

P9. *I honestly do not think that our students read the rubrics or the comments or anything or take notes unless you actually get them into your office and give them a lecture about what they are doing ...one to one. But I am sure they don't actually look at the rubric. They just look at the score. The grade is what gets them, and they will come after you and they will ask you "Why did you grade me like that?" They don't really read the rubric. They just want to know "Why did you give me this grade?"*

P 3. *Honestly, I had to show my students [how to use e-rubrics] multiple times. I teach them rubrics in Word document form and ... I had them click on their own assessments after I put*

them on Turnitin and check the rubrics themselves and I did a dummy practice session in class just to have them see how much function and format were worth and how much content and grammar were worth. I think that did help them but as in the past I asked some of my students they never even saw, they never even read, they never even tried to. Honestly, I don't think they pay attention to rubrics at all until they have to.

Students do not read the rubric and instructors spend a lot of time and effort to show them how to access the e-rubric and even give a verbal recap of the feedback to the student, so that they can take some action to improve their essay. They also spend quite an amount of time to construct a successful e-rubric to be able to communicate the course expectations to their students.

3.1.2. Research question 2

The instructors were asked how e-rubrics compare to traditional, paper-based rubrics. The analysis of the data regarding this research question revealed three main themes: reducing the workload in grading, providing instant feedback to students, and safe record keeping. Besides these advantages, two instructors also mentioned that paper-based rubrics are as efficient as e-rubrics, although e-rubrics' superiority in providing immediate feedback to students is undisputed, as shown in Table 4 below.

Table 4. *E-rubrics vs. paper rubrics.*

Advantages of e-rubrics over paper rubrics	Frequency
Reducing the workload in grading	9
Making feedback instantly accessible to students	5
Safe record keeping	3

All instructors agreed that using e-rubrics reduces the time and effort spent on grading and scoring student essays. Reducing the workload was defined as making grading easier, quicker, and providing flexibility in adapting the same rubric for other assignments and courses. Direct quotes from instructors to support their viewpoints are provided below:

P5. *They are faster in the sense that if you got the essay, you got the rubric. It is part of the same sort of interface. Therefore, I think you can mark or grade the essay more quickly with more clarity because you can see the rubric and you can see the essay at the same time.*

P8. *They are much faster and much easier. I like the use of Turnitin. It makes it a lot easier. I like the fact that the rubrics are attached, and quick marks are easy to use. You can give students instant feedback on their issues and Turnitin seems to be able to track students' progress or similarities in the issues that students have. The report that Turnitin can generate would help me to assess a class's level and the issues they have and what we need to do about it and how to improve. The descriptors on the attached rubrics are very clear and easy to use, and I like the way they can change as you move across each section of the rubric.*

Providing immediate feedback to students was perceived to be another superiority of e-rubrics. Instant feedback means students are able to act on the feedback quicker and will have more time to revise their essay.

P3. *E-rubrics are much easier to get through. The weighting of each and every single category is kind of difficult because…for example, content could be scored higher than grammar and mechanics. I think e-rubrics are beneficial not just for me as a teacher to make the grading process much easier but also easier for the student to see how they can achieve let's say a higher grade in each and every single assessment and what to avoid.*

P6. *The only thing is that it is just easier to select the scale and they are more accessible to students because they would see exactly where they would fall on every category so I would say they are more accessible and they are more flexible but in terms of effectiveness or efficiency, I would say they are same as paper rubrics.*

Safer record-keeping particularly preventing papers getting lost is another advantage that the e-rubrics offer as opposed to paper rubrics. Students and instructors alike will not have to file essays and their assessment in a folder; they will always be online.

P7. *E-rubrics are easier to use, of course. You don't need to print them out beforehand and prepare them. You know all this; they are there all the time so they cannot be lost. The students can refer to them. This is a very good thing because you cannot lose this. Also, e-rubrics are easier so basically you are marking things online. You don't have to go through the hassle of printing and then writing on paper.*

P1. *Basically, paper rubrics and e-rubrics are about the same. But they, e-rubrics, have two advantages. Students can't lose them, so they are always there and they are available. For conferencing I don't have to go hunting through anything. I just open up Turnitin and find the student's name and there is the rubric attached.*

3.1.4. *Research question 4*

The last question in the qualitative section is what instructors think about the impact of rubrics on student learning. This question did not elicit as straightforward responses as the previous ones, because not all instructors were convinced that e-rubrics have a significant impact on students' writing performance and their responses generally ranged in a continuum of "rubrics have no impact" to "rubrics have a limited impact on student learning".

Therefore, instead of creating categories, and displaying the frequency of responses, only direct quotations were given below to reflect the participants' observations from pessimistic to optimistic in regards to the positive impact of rubrics on student learning.

P9. *So, I haven't seen any improvement or any awareness in my students. Not from the rubrics. They don't really read the rubric. They just want to know "Why did you give me this grade?"*

P1 *I have to be honest, no. Well…maybe 5 percent. The students that use them, most of them don't even bother. I would say only 5 percent of the students, but I would still argue for them because it protects us especially if a student wants to do a grade appeal.*

P4. *To an extent, yes, but not completely... This is because students won't necessarily read the whole thing, which means they won't clearly understand what is expected of them.*

P5. *Yes, for the students who care, I think I could say there has been an improvement. I wouldn't say a huge improvement, but I would say there is an improvement because they are more conscious of what they need to be putting into each paragraph and they are more conscious of the necessary structures and because they have a better awareness of the grades they will get for individual parts of an essay.*

P6. *I believe students who care do benefit from rubrics. First, because when they look at the rubric, they… before even working on their assignments… they know exactly how they are being graded and they can get an idea about what's expected for them to get an A. I think it lays out the expectations very clearly and they usually post it with the assignment sheet and they say this is what you are supposed to do to get an A and I sometimes use it for in-class activities like during the drafting stage I have students grade each other's papers using the grading rubric so that helps them see things from my perspective and understand what exactly I expect so this is very helpful in terms of working on the assignment.*

As the responses display, writing faculty think students benefit from rubrics only when they care and when the course instructors spend a substantial amount of class time or one-on-one

tutoring time to explain how the e-rubrics works. This could be interpreted as they do not expect rubrics to work a miracle and assume their students learn from e-rubrics on their own.

3.1.5. Research question 5

In the quantitative part of the study, MFRM was applied to the raters' data to examine if there is any potential link between their beliefs towards using e-rubric and their rating behaviour. In a rating situation, raters are expected to perform similar levels of severity/leniency which is not always possible. The severity measures by FACETS are used to analyze if the inconsistencies among raters' severity levels are significant. According to the results on the relevant output (Table 5), P7 was found the most severe in ratings with a measure of 2.62 while P6 was found the most lenient with a measure of -2.40. The other raters' severity measure varied within a small range -.43 and .65, indicating that those four instructors were not substantially different in terms of the levels of severity they exercised.

Table 5. *Rater-related statistics.*

Rater ID	Measure[*]	S.E.[***]	MnSq[**]
P1	.60	.07	2.77
P2	.13	.09	.61
P5	.64	.17	.51
P6	-2.40	.08	.23
P7	2.62	.12	.54
P9	-.43	.05	.72

[*]Measure: Rater's severity measure in log-odd units. Higher value indicates a more severe rater and vice versa.
[**]MnSq: Outlier sensitive fit statistics value for each rater.
[***]Standard error of the estimated measure

Another statistic analyzed in this part is the rater separation index. Rater separation index value shows the number of different strata of severity in the raters. Since the raters are expected to perform a similar level of severity in theory, the expected value of this statistics is 0. In the study, the separation index value found 5.23 which indicates that within 6 raters, there were about six statistically distinct strata of severity. Reliability of rater separation is also checked to understand rater severity. It shows the degree that the raters can be differentiated in terms of their severity level. Similar to the rater separation index, a value of 0 is expected for this statistic. This value was found .29 which indicates the raters were differentiated fairly in terms of levels of severity they exercised. Lastly, fixed (all same) chi-square and its significance value which test if the raters significantly differ in their levels of severity were checked. According to the results, the chi-square value of 14.2 with a significance value smaller than .01 indicates that the severity measures for the raters were not all the same, after allowing for measurement error.

3.1.6. Research question 6

In addition to the severity measures that outline systematic rating behaviors, we analyzed rater fit statistics, particularly mean-square infit and outfit indices which show the degree of unexpected ratings by a rater summarized over examinees. Fit statistics allows researchers to examine if any rater effect exists in the ratings for some examinees or items/tasks. Based on the results, only P1 had a mean-square outfit value (2.77) that implies a significant inconsistency in ratings. This finding suggests that the particular rater may have adopted an *idiosyncratic rating style* (Eckes, 2011) since mean-square outfit statistics is sensitive to outliers in ratings.

4. DISCUSSION

Our findings on the first research question showed that writing teachers benefit greatly from e-rubrics. They specifically expressed that using e-rubrics makes grading transparent, reduces student complaints and grade appeals, facilitates grading and helps standardization among raters. Recently e-rubrics have also been analyzed as computer assisted grading rubrics and LMSs (such as Blackboard and Moodle) integrated rubrics, however there are not many studies conducted on teachers' perceptions towards electronic rubrics, probably due to its novelty.

In general, rubrics are perceived positively by academic faculty and teachers. Instructors have a positive attitude towards rubrics as a handy evaluative and instructional tool for nurturing students' learning, self-assessment, and self-regulation (Sharma, 2019); and most teachers agree that rubrics act as a guideline for students to know what the criteria are in order to get a good grade and that they became more consistent in grading since they started using rubrics (Qasim & Qasim, 2015). Regarding e-rubrics, Raddawi and Biliközen (2018) express that the instructors think using e-rubrics helps them with record keeping, saves time and energy in grading and giving feedback and provides objective assessment of student essays. Atkinson and Lim (2013) also suggest e-rubrics are efficient for grading and giving feedback, they provide detailed guidance to students, and promote standardization.

An interesting finding that stands out in our study is the benefit of rubrics in reducing student complaints or grade appeals. Although we have not been able to see a similar finding in other recent studies, Fulbright (2018) suggests that rubrics add transparency to grading, which is important when explaining to disgruntled students that they were not given a certain score because the instructor did not like them, but because they omitted one or more components of the required criteria of the assignment. Fulbright (2018) asserts that rubrics give faculty the needed documentation of objective assessment that is essential for grade appeals and even in court. The fact that this benefit of rubrics has been frequently expressed by the instructors in this institution may be related to the culture of the institution. If instructors are getting a high number of grade appeals and complaints in this institution, the use of rubrics may have been adopted as a defense mechanism.

The second finding was about the limitations of e-rubrics and the writing faculty expressed that they face scoring issues such as rubrics not always providing them with the clarity and flexibility they need to score papers. Also, some problems confronted by students such as not being able to access or not reading the rubric and sometimes even not understanding the rubric are frequently mentioned by the instructors interviewed.

Raddawi and Biliközen's (2018) study elicited similar responses from writing instructors with regards to the rigidity of rubrics. Lack of flexibility that they experience while grading essays with a rubric seems to be restricting teachers' freedom. In their study technical problems were expressed by the writing faculty, however in our study this was attributed as a student challenge due to their lack of IT skills. The fact that no teacher has mentioned any technical problems in regard to the use of e-rubrics is a sign that they have all mastered the use of e-rubrics in the institution.

Another similar finding comes from a study conducted on student perception on the e-rubrics. In Atkinson and Lim's study (2013) when students were asked to make suggestions to improve the rubrics their professors use, some of them said "less ambiguity and clearer requirements and relevant depth are needed. It was hard to customize the task because the structure was already defined." Some students even mentioned they would like to see more freedom included in the assignment task. A similar study analyzing university students' perceptions towards using rubrics showed that despite the many positive aspects of using rubrics, a small percentage of students indicated that the rubric lacked clarification because it contained standard feedback

for everybody (Raposo-Rivas & Gallego-Arufat, 2016) Sharma (2019) also express that especially low achieving students do not show any interest in practices like using rubrics for self-assessment unless they are specifically trained to do so, which also supports our interviewees' responses in regards to students not reading the rubrics and not accessing the rubrics (unless the professors spend a long session on how to access the rubrics online).

The third finding of the study was about how e-rubrics compare to traditional rubrics in the writing faculty's opinion. This research question revealed that although they are more or less the same in content and structure, e-rubrics are superior to paper rubrics because they reduce the workload in grading, make feedback instantly accessible to students and help instructors with record keeping.

A study comparing the use of computer-assisted grading rubrics to other grading methods and their results suggested that the computer-assisted grading rubrics were almost 200% faster than traditional hand grading without rubrics, more than 300% faster than hand grading with rubrics, and nearly 350% faster than typing the feedback into a Blackboard or Moodle (Anglin et al., 2008). Atkinson and Lim (2013) also found out that a key benefit of e-rubrics was around 40% reduction in marking time. Indeed, rubrics embedded in a learning management system (in our study, Moodle) not only make grading faster, but also record keeping much easier as they facilitate student submissions, and help faculty track details such as student names, uploaded files, similarity rates and the time of submission. Besides, automated calculations enabled by the LMSs ensure speed and accuracy. Results are available for general feedback to students, and for examination and auditing by other stakeholders.

The last finding of the qualitative analysis is about the perceptions of the writing faculty on whether rubrics have any impact on student learning. The interviewees in this study were slightly hesitant to make comments about the positive impact of rubrics on student writing; instead, they tended to see some improvement in students who really cared about self-assessment and who made the effort to go through the rubric and attended conference hours organized by the professor to get more feedback along with the extra explanation in the rubric.

This cautious optimism about rubrics could be also found in some studies done previously. Reddy and Andrade (2010) express that there is a striking difference between students' and instructors' perceptions of rubric use. College students value rubrics because they clarify the targets for their work, allow them to regulate their progress, and make grades transparent and fair. While students referred to rubrics as tools serving the purposes of learning and achievement, instructors focused almost solely on the role of a rubric as a tool to assign grades quickly, objectively, and accurately. This could be the underlying reason for the interviewees in this study to have doubts about the value of rubrics as a teaching tool because they value rubrics as something that makes their grading easier and reduces student complaints.

A rubric skeptic, Krane (2018) also suggests rubrics do not teach students how to write and foster deep learning. She conducted a study with her 88 students and 69% agreed or strongly agreed that rubrics should always be given with writing assignments; they liked discussing rubrics in their classes and referring to them when working on assignments. 86% noted rubrics helped them to understand what the professor wants. 83% noted rubrics helped them to understand assignment criteria, and 74% noted that rubrics helped them "to know what they can do to get a better grade". However, when the students were asked whether rubrics helped them to improve their writing in general, only 21% agreed.

Rubrics are assignment specific; therefore, they highlight what will be assessed in a given assignment. Therefore, they work well in the short run, especially when students need guidance and a roadmap to get a good grade. However, as the interviewees' responses suggested, this roadmap does not necessarily help them develop problem solving skills that would improve

them as writers in the long run. Not all writing instructors think rubrics work miracles for students who do not make the effort to develop their writing skills. Those who care about rubrics are probably conscientious students who would strive to develop their writing without a rubric, anyway.

Quantitative results overlapped with qualitative findings, revealed by the raters' self-report. For example, the finding related to research question five was about the raters' severity and leniency. Although there is no consensus on personal and situational determinants of rater severity effect, professional experience is commonly cited as a factor leading to severity effect (Eckes, 2011). In our study, P7 was the most severe and P6 was the most lenient rater, as per our analysis. This could partially explain the most severe rater's (P7) rating behavior who has only 6 months of experience with e-rubrics (s/he has only 6 months of experience in the institution, as well). This may be the reason why as the least experienced rater, P7 may have tended to notice even the smallest errors in the name of using e-rubric carefully and turned out to be a harsh grader.

Another interesting finding in the quantitative study was P1 turning out to be an inconsistent rater, as suggested by FACETS. Inconsistencies in grading stem from several factors related to the problem being graded, the individual grader, the time of day, the grader's level of fatigue, and the grader's overall experience. Graders are also affected by their general values and beliefs about grading, such as values of non-achievement factors, like effort, and perceptions that grades function as rewards or punishments (Hicks & Diefes-Dux, 2017).

Reddy and Andrade (2010) emphasize the striking difference between students' and instructors' perceptions of rubric use; while students referred to rubrics as a learning tool, instructors focused on the role of a rubric as a tool to assign grades quickly, objectively, and accurately. In P1's case, this purpose may have been solely to "stuff the grades" s/he is giving and justifying this grade in case s/he gets a complaint or a grade appeal. This "limited conception of the purpose of a rubric" (Reddy & Andrade, 2010) might have contributed to their unwillingness to use the rubric properly and consequently, assign their grades randomly (p.439).

5. CONCLUSION & RECOMMENDATIONS

The findings of this study reveal that although freshman composition instructors seem to be enjoying the advantages of e-rubrics, they differ in their rating behavior while implementing the standard e-rubric in assessing writing. Rater severity and inconsistency may cause dissatisfaction and create a sense of unfairness among the students of such instructors; therefore, instructors should be made aware of their rating behavior, so that they can avoid repeating them. Instructors may require some clarification on how to interpret some items in the rubric and they may need to be convinced about the educational value of rubrics as well as the evaluation or justification tool as they have been using it.

The most common way of fulfilling this goal is training sessions, where instructors are introduced to a set of criteria and then they are asked to rate essays based on those criteria. The results show whether and to what extent they are on the same page as other raters and therefore interpret the rating criteria similarly. Raters who still show severity/leniency and/ or inconsistency may take additional rater training sessions to prevent such rater effects in the future. Organizing a norming session before every essay assessment is also an efficient means for departmental standardization. Consequently, the results of this study are expected to help this writing department to be more standardized in its ratings.

Writing assessment is a challenging job and writing teachers invest a lot of time and effort into helping their students improve their writing. Their rating behaviour may differ but writing scholars are interested in working collaboratively to discover the most effective method of assessing writing. Therefore, the implications of this study should not be seen as limited to the

particular institution or the region where the study was held. Even though the findings cannot be generalized, they bring out some serious considerations concerning the application of rubrics as writing assessment tools in an EFL (English as a foreign language) context. This study has implications for many rater-mediated language assessment situations, particularly in small-scale academic programs.

Last but not the least, a recommendation for researchers could be regarding the scope of such a study, which could be further enhanced by collecting data from students regarding their perceptions of e-rubric as an instruction and assessment tool and even assessing their rating behavior while they are using the rubric in peer grading.

5.1. Limitations

This study had several limitations. For the qualitative part, it should be noted that the faculty members probably had the standard e-rubric in mind that was in place at the time of the interview, so their perceptions and experiences were predominantly shaped by that particular e-rubric they were required to use in the department.

For the quantitative part, the limitations are the e-rubric and grading scale that are in use as well as the genre (research paper) that the students submitted as their assignment. It should also be added as a limitation that this was the second draft of the assignment, which means the students revised their first drafts based on the feedback they received from their instructor and submitted an improved version as a second draft. This may have positively skewed their grades.

Acknowledgments

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the author(s).

Authorship Contribution Statement

Inan Deniz Erguvan: Data Collection, Resources, Qualitative Analysis, and Writing the Draft. **Beyza Aksu Dünya:** Software, Quantitative Data Analysis and Reporting, Editing the Draft.

ORCID

Inan Deniz ERGUVAN 🆔 https://orcid.org/0000-0001-8713-2935
Beyza AKSU DÜNYA 🆔 https://orcid.org/0000-0003-4994-1429

6. REFERENCES

Allen, D., & Tanner, K. (2006). Rubrics: tools for making learning goals and evaluation criteria explicit for both teachers and learners. *CBE Life Sciences Education*, *5*(3), 197-203. https://doi.org/10.1187/cbe.06-06-0168

Andrade, H. G. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57(5), 13-18.

Andrade, H. G. (2005). Teaching with rubrics: The good, the bad, and the ugly. *College Teaching*, *53*(1), 27. https://doi.org/10.3200/CTCH.53.1.27-31

Anglin, L. Anglin, K., Schumann, P.L., & Kalinski, J. A. (2008). Improving the efficiency and effectiveness of grading through the use of computer-assisted grading rubrics. *Decision Sciences: Journal of Innovative Education, 6*(1), 51-73. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-4609.2007.00153.x

Arter, J., & McTighe, J. (2001). Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance. Corwin Press.

Atkinson, D. & Lim, S.L. (2013). Improving assessment processes in higher education: Student and teacher perceptions of the effectiveness of a rubric embedded in a LMS. *Australasian Journal of Educational Technology*, *29*(5), 651-666. https://doi.org/10.14742/ajet.526

Broad, B. (2000). Pulling your hair out: "Crises of standardization in communal writing assessment". *Research in the Teaching of English, 35*(2), 213-260. www.jstor.org/stable/40171515

Brookhart, S. M. (2013). *How to Create and Use Rubrics for Formative Assessment and Grading*. ASCD.

Brookhart, S. M. (2018). Appropriate criteria: Key to effective rubrics. *Frontiers in Education*, *3*. https://www.frontiersin.org/articles/10.3389/feduc.2018.00022/full

Brookhart, S. M., & Chen, F. (2015). The quality and effectiveness of descriptive rubrics. *Educational Review, 67*(3), 343-368. https://doi.org/10.1080/00131911.2014.929565

Brown, J. (2001). *Using surveys in language programs.* Cambridge University Press.

Carr, N. T. (2000). A comparison of the effects of analytic and holistic rating scale types in the context of composition tests. *Issues in Applied Linguistics*, *11*(2), 207-241 https://escholarship.org/uc/item/4dw4z8rt

Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among five approaches* (2nd ed.). SAGE.

Creswell, J. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Pearson Education.

Creswell, J. W., & Clark, V. L. (2011). Choosing a mixed methods design. In *Designing and conducting mixed methods research* (3rd ed., pp. 53, 106). SAGE.

Dobria, L. (2011). *Longitudinal rater modeling with splines.* (Publication no. 3472389) [Doctoral dissertation, University of Illinois at Chicago]. ProQuest digital dissertations.

Eckes, T. (2011). *Introduction to many-facet Rasch measurement: Analyzing and evaluating rater-mediated assessments.* Peter Lang.

Freeman, S., & Parks, J. W. (2010). How accurate is peer grading? *CBE Life Sciences Education*, *9*(4), 482-488. https://doi.org/10.1187/cbe.10-03-0017

Fulbright, S. (2018, October 18). *Using rubrics as a defense against grade appeals.* Faculty Focus.https://www.facultyfocus.com/articles/course-design-ideas/rubrics-as-a-defense-against-grade-appeals/

Garcia-Ros, R. (2011). Analysis and validation of a rubric to assess oral presentation skills in university contexts. *Electronic Journal of Research in Educational Psychology*, *9*(3), 1043-1062.

Hafner, J.C., & Hafner, P.M. (2003). Quantitative analysis of the rubric as an assessment tool: An empirical study of student peer-group rating. *International Journal of Science Education*, *25*(12), 1509-1528. https://doi.org/10.1080/0950069022000038268

Hicks, N., & Diefes-Dux, H. (2017). Grader consistency in using standards-based rubrics. 2017 ASEE Annual Conference & Exposition Proceedings. https://doi.org/10.18260/1-2--28416

Iramaneerat, C., & Yudkowsky, R. (2007). Rater errors in a clinical skills assessment of medical students. *Evaluation & the Health Professions*, *30*(3), 266-283. https://doi.org/10.1177/0163278707304040

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

Kirwin, J., & DiVall, M. (2015, October). Using electronic rubrics to produce actionable assessment data in a skills-based course [Conference session]. 2015 Assessment Institute in Indianapolis, Indianapolis. https://assessmentinstitute.iupui.edu/

Kohn, A. (2006). Speaking my mind: The trouble with rubrics. *English Journal*, *95*(4), 12-15. https://doi.org/10.2307/30047080

Krane, D. (2018, August 30). *Guest post: What students see in rubrics.* Inside HigherEd. https://www.insidehighered.com/blogs/just-visiting/guest-post-what-students-see rubrics

Linacre, J. M. (2019). *A user's guide to FACETS: Rasch-model computer programs.* https://www.winsteps.com/tutorials.htm

Linacre, J. M. (2020). *FACETS* (Version 3.83.2). https://www.winsteps.com/winbuy.htm

Linacre, J. M. (2002c). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16,* 878. https://www.rasch.org/rmt/rmt162f.htm

Martínez, D., Cebrián, D., & Cebrián, M. (2016). Assessment of teaching skills with e-Rubrics in Master of Teacher Training. *Journal for Educators, Teachers and Trainers, 7*(2). 120-141. https://jett.labosfor.com/article_855.html

Mertler, C. A. (2009). *Action research: Teachers as researchers in the classroom* (2nd ed.). SAGE.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook.* SAGE.

Myford, C. M., & Wolfe, E. W. (2003). Detecting and measuring rater effects using many-facet Rasch measurement. *Journal of Applied Measurement*, 5(2), 189-223.

Myford, C. M., & Dobria, L. (2012). *FACETS introductory workshop tutorial.* University of Illinois at Chicago.

Nordrum, L., Evans, K., & Gustafsson, M. (2013). Comparing student learning experiences of in-text commentary and rubric-articulated feedback: Strategies for formative assessment. *Assessment & Evaluation in Higher Education*, *38*, 919-940. https://doi.org/10.1080/02602938.2012.758229

Qasim, A., & Qasim, Z. (2015). Using Rubrics to Assess Writing: Pros and Cons in Pakistani Teachers' Opinions. *Journal of Literature, Languages and Linguistics*, *16*, 51-58. https://www.researchgate.net/publication/285815750

Raddawi, R., & Bilikozen, N. (2018). ELT professors' perspectives on the use of E-rubrics in an academic writing class in a University in the UAE. Assessing EFL Writing in the 21st Century Arab World, 221-260. https://doi.org/10.1007/978-3-319-64104-1_9

Radnor, H. A. (2002). *Researching your professional practice: Doing interpretive research*. Open University Press.

Raposo-Rivas, M., & Gallego-Arrufat, M.J. (2016). University students' perceptions of electronic rubric-based assessment. *Digital Education Review, 30*, 220-233. http://greav.ub.edu/der/

Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, *35*(4), 435-448. https://doi.org/10.1080/02602930902862859

Rivasy, M. R., De La Serna, M. C., & Martínez-Figueira, E. (2014). Electronic rubrics to assess competencies in ICT subjects. *European Educational Research Journal*, 13(5), 584-594. https://doi.org/10.2304/eerj.2014.13.5.584

Saal, F. E., Downey, R. G., & Lahey, M. A. (1980). Rating the ratings: Assessing the psychometric quality of rating data. *Psychological Bulletin*, *88*(2), 413-428.

Sadler P.M., & Good E. (2006). The impact of self-and peer-grading on student learning. *Educational Assessment*, *11*(1), 1-31. https://doi.org/10.1207/s15326977ea1101_1

Sharma, V. (2019). Teacher perspicacity to using rubrics in students' EFL learning and assessment. *Journal of English Language Teaching and Applied Linguistics*, *1*(1), 16-31. https://www.researchgate.net/publication/337771674

Silverman, D. (2000). *Doing qualitative research: A practical handbook.* SAGE.

Steffens, K. (2014). E-rubrics to facilitate self-regulated learning. REDU.*Revista de Docencia Universitaria*, 12 (1), 11-12. https://doi.org/10.4995/redu.2014.6417

Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education: Principles, Policy & Practice, 14* (3), 281-294. https://doi.org/10.1080/09695940701591867

Wilson, M. (2007). Why I won't be using rubrics to respond to students' writing. *English Journal, 96*(4), 62-66. https://doi.org/10.2307/30047167

7. APPENDIX

The rubric for the research paper

	Excellent 100	Very good 90	Good 80	Average 70	Inadequate 60	Poor 10
Research Elements (Sources & Quotations) **25%**	Paper uses 6-7 quality sources (mostly scholarly), and provides quotations from a variety of sources in <u>every</u> body paragraph. The quotations support the topic sentences well. Support and evidence are expressed in the writer's words (paraphrased & summarized).	Paper uses min 5, preferably 5-6 quality sources (popular & scholarly), and provides quotations from a variety of sources in <u>every</u> body paragraph. The quotations support the topic sentences well. Support and evidence are usually expressed in the writer's words.	Paper uses 5 good sources, and provides quotations from a variety of sources in <u>most</u> body paragraphs. The quotations generally support and develop the topic sentences. Support and evidence are mostly expressed in the writer's words, but some direct quotes are unnecessary.	Paper uses 4-5 sources, but 1 -2 may be weak or not academic enough. The evidence may be irrelevant/ weak in 1-2 body paragraphs. Support and evidence are not always expressed in the writer's words. Word count may be low due to lack of sources & quotations.	Paper uses and provides quotations from less than 4 sources. The evidence is weak and irrelevant, does not develop the thesis. Support and evidence are not expressed in the writer's words. Word count is below 1000.	There is no indication of research. No sources have been used. Very low word count, or high similarity rate. Too many direct quotations.
Organization & Connectors **25%**	Introduction is interesting with detailed background and a clear thesis statement. Topic sentences introduce the arguments, body paragraphs fully explore the topic and present information in a logical order. Counter argument has a strong refutation. Conclusion restates the thesis and contains original opinions. Effective and varied transitions link all ideas.	Introduction gives good background and contains a clear thesis statement. Topic sentences introduce the arguments; body paragraphs explore the topic and present information in a logical order. Counter argument has a relevant refutation. Conclusion restates the thesis and contains opinions. Transitions link all ideas.	Introduction gives some background and contains a thesis statement. Topic sentences exist in every body paragraph. The arguments in body explore the topic and present information in an acceptable order. There is a counter argument with some refutation. Conclusion restates the thesis and/or offers a comment. Transitions link most ideas.	There is an underlying organization, but paragraph division may not always be efficient. Introduction contains a thesis statement. The body explores the topic and presents information, but not always clear or logical. Counter argument and/or refutation is weak or refutations is nonexistent. Some transitions are used, but more are needed.	The paper is generally hard to follow. The writing lacks strong organization and it may also lack a clear thesis statement. The body presents some support, but not all relevant. Transitions may be used inconsistently or may be lacking.	There is no clear paragraphing. Introduction, body and conclusion paragraphs are not clearly divided. Lack of transitions impedes fluency.

Grammar / Mechanics/ Spelling/ **25%**	There are 1-2 minor errors in grammatical accuracy. Spelling and punctuation may contain 1-2 typos. Word choice is appropriate for an academic research paper. Complex sentences (noun, adverb, adjective clauses) are frequently used, without errors.	There are 3-4 minor errors in grammatical accuracy. Spelling and punctuation may contain 1-2 errors. Word choice is generally good for an academic research paper. Compound & complex sentences are frequently used, with minor errors.	Up to 5 errors in grammatical accuracy, spelling, or punctuation may exist. The use of academic words is acceptable. Sentence variety is not as expected, complex sentences may contain max 3 errors.	There are (max 10) errors in grammatical accuracy; some may detract from the meaning. There is not enough evidence of academic vocabulary. Preference for simple sentences, rather than complex.	There are more than 10 grammar, spelling & punctuation errors. Word choice is incorrect or inappropriate in most places. Writing is choppy, with many awkward or unclear passages.	Very poor use of English with no sense of correct grammar. Google translation or synonym finder may have been used.
APA & Formatting **25%**	Consistently uses accurate in-text citations and has a flawless Reference page (alphabetical, double-spaced, in hanging indent format). All sources cited in the essay are listed on the Reference page. Entire essay is double-spaced using Times New Roman font with 1-inch margins. Student's name, instructor's name, course, date appear as per APA guidelines. 1 or 2 minor errors may exist in overall formatting. (Essays with no title cannot get excellent in this category)	Max 3 minor errors exist in in-text citations and/or Reference page. Essay is notably lacking in 1-2 items in the Excellent category (For example, double spacing / margins or etc.)	In-text citations and/or Reference page have 5 major errors. Essay is notably lacking in 3-4 items in the Excellent category (For example, double spacing / margins or header etc.).	The references page may lack 1-2 sources or contain formatting errors. There are 5-6 errors (page number, font, spacing) and major (in text citations & references) in APA style.	Most formatting is incorrect or inconsistent. APA is not adhered to in in-text citations and references. There are more than 5 major errors.	No references page submitted or lists some sources with incorrect style. No formatting style has been followed throughout the paper. Citations don't exist.

Note: The large gap between Inadequate (60) and Poor (10) intends to serve the purpose of differentiating D students from F students. D students are still considered passing and with a little more effort can get a C. However, F students perform quite below the expected standards, therefore they should make greater effort to pass the assignment and eventually the course. It should also be noted that as per the feedback from course instructors, rubrics are revised every year.