



THE AFFORDANCES OF RUBRICS IN L2 WRITING IN HIGHER EDUCATION

A new approach to enhancing writing conventions

AITOR GARCÉS-MANZANERA
Universidad de Murcia, España

KEYWORDS

*rubrics
L2 writing
Higher Education
feedback
writing genre
EFL writing
writing assessment*

ABSTRACT

The use of diverse techniques for the evaluation of writing tasks in English as a Foreign Language (EFL) has made its way into the EFL classroom in order to facilitate both the teachers' task and the L2 students' comprehension. Thus, the aim of this paper is to explore how undergraduate students may be trained in the use of rubrics, an ecologically valid feedback technique, and how they might assess sample writing tasks. This way, we will observe how able they are to identify dissimilarities and consistencies in these writing tasks on the basis of a specific genre as an article.

PALABRAS CLAVE

*rúbricas
escritura en L2
Educación Superior
retroalimentación
género de escritura
escritura en inglés
evaluación de la escritura*

RESUMEN

El uso de técnicas diversas para la evaluación de tareas escritas en Inglés como Lengua Extranjera ha hecho su entrada en la clase de lengua extranjera con el objetivo de facilitar la tarea de la profesora y los estudiantes. Así, el objetivo de este estudio es explorar como los estudiantes de grado pueden ser entrenados en el uso de rúbricas, una técnica de feedback ecológicamente válida, y como pueden evaluar tareas de writing de ejemplo. De esta forma, se observará su capacidad para identificar similitudes y diferencias en estas tareas de writing tomando como base un género específico como el artículo.

Recibido: 06/ 07 / 2022

Aceptado: 12/ 09 / 2022

1. Introduction

Novel practices in the second language (L2) writing classroom have given rise to a series of approaches in Higher Education, leading to the incorporation of methodological tool to ensure that assessment is fair and clear.

In L2 writing, extensive scholarly interest has emerged out of the necessity to provide students with sound feedback. A newcomer in the Second Language Acquisition (SLA) strand of research (Manchón, 2009, 2011), the introduction of diverse feedback techniques has broadened the scope of action. As a result, methods for evaluating texts have become relevant. Among these many techniques, the use of rubrics has been a handy ally of EFL teachers, especially in Higher Education, in which the number of students makes it unbearable to correct to detail. Whilst the scholarly literature has accumulated evidence regarding the use of rubrics from different angles, e.g. peer-assessment, teacher-assessment or self-assessment, to our knowledge, there scarce studies exploring how undergraduate students are trained to exploit rubrics for their L2 writing improvement, and how they apperceive L2 writing assessment. Thus, the present study has different objectives: (1) determine how rightful L2 students are when assessing writing tasks, (2) the extent to which they understand the writing conventions of a specific genre (e.g. articles, reviews, among others), and (3) whether they identify correctly pieces of writing containing deliberate errors.

2. Writing as an additional language: writing-to-learn

In the recent decades, writing in an additional language (L2) has received more scholarly attention as it has been observed in the amount of research produced (see Leki *et al.*, 2008; Manchón & Polio, 2021, for a comprehensive review). In essence, the study of writing in an L2 has been dealt with from very varied perspectives, one of which has been the assessment of the written product through a number of feedback techniques (see Bitchener & Storch, 2016; Manchón, 2009).

As a result of this increasing scholarly interest in L2 writing, studies exploring both theoretical and empirical perspectives began to generate growing interest in the recent SLA studies as well as Instructed Second Language Acquisition scientific literature (see Bitchener & Storch, 2016; Manchón & Vasylets, 2019). In this regard, Manchón (2011) established a dual distinction in the study of writing: (a) the *learning-to-write* dimension, and (b) the *writing-to-learn* dimension. In the case of the *learning-to-write* dimension, its focus is placed on how a learner might gain knowledge about the conventions or style of a particular writing genre (e.g. academic writing), and the manner in which this skill may be developed in an L2. As regards (b), that is, the *writing-to-learn* dimension, it involves exploring and understanding how the act of writing and of providing feedback may contribute to creating the adequate environment for language learning. Manchón & Vasylets (2019) indicate that there are three characteristics that make writing as an ideal site for language learning: (i) the availability of time while writing, and subsequently, while processing feedback; (ii) the permanence of writing, and (iii) the force triggered by the problem-solving nature of writing and the depth of processing while processing feedback (Leow, 2020).

Additionally, there have been manifold approaches to writing instruction both in the first language (L1) and the L2. Among these, the process approach (Hyland, 2016) has been highlighted given its focus on the different stages, that is, planning, drafting and revising. In this regard, the final process of writing has been considered as a fully individual, yet cognitive activity. When L2 writers are editing or revising their texts, they engage in a series of operations which are destined to examine the text from a variety of perspectives. In this equation, the role of the teacher is central as a number of aspects have to be highlighted: for instance, grammar, vocabulary but also organization and communicative achievement.

Texts in an L2 are, thus, subject to a more in-depth observation which is mainly related to the concept of *genre*. Within the three schools of thought which have dealt with the theoretical relevance of genre, Systemic Functional Linguistics understands genre as mainly a social practice (Tardy, 2012, as cited in Rock, 2021). In essence, this scholarly current views genre as a goal-oriented social activity in which a series of sub-constructs are included (Martin, 2002, 1999). Texts – be it in the L1 or L2 – are usually subject to a specific text type wherein genre comes into play. Among the characteristics of genre, *register* is a central one since it involves a series of lexicogrammatical choices (Byrnes, 2012) highly dependent on: (i) the field, (ii) the tenor, and (iii) the mode. The field helps the L2 writers to materialize the perspective to take in the written text, for instance, when a report is written the social conventions are different from writing an essay. In the case of (ii), the tenor characterizes the social interaction between the text and the individuals at which this text is going to be addressed, and finally (iii) involves the mode as regards whether the message is communicated verbally or through a written medium.

Connecting the role of writing as a site for language learning and the approaches to writing instruction which have incorporated genre-based perspectives, our pivotal aim in this study is to shed light upon L2 undergraduate learners' real perceptions of genre on the basis of second language writing pedagogy. In light of this, the assessment of writing has been traditionally associated as part of the teachers' responsibilities since it is a fundamental teaching task (Ferris & Hedgcock, 2014). The spectrum of assessment possibilities in the L2 is varied, but in the context of L2 writing, assessment is generally restricted to either (1) feedback or (2) the use of

rubrics. The assessment of writing in the form of feedback in the final written product has received considerable scholarly attention in recent decades, especially as written *corrective* feedback (see Bitchener & Storch, 2016; Polio, 2012). This type of writing assessment involves a more dedicated, time-consuming approach to correcting L2 writing, and it has been demonstrated to be particularly effective in its various forms (e.g. focused written corrective feedback or direct written corrective feedback). Nevertheless, it is a more teacher-centered type of feedback which involves corrections at a micro-textual level. As a result, Ferris (2014) indicated that rubrics (2) could be another more efficient way of providing feedback. One of the advantages of rubrics against traditional feedback is the explanations that L2 writers may receive in regard to their written text, not only focusing on the lexicogrammatical aspects but also on the characteristics of the genre. In fact, while traditional feedback (e.g. underlining, indicating the correction) is generally restricted to the teacher's sphere of action given the expertise required, the use of rubrics may be extended as an instructional practice to implement in the L2 writing classroom.

Building on this statement, our study revolves around the use of rubrics in L2 writing assessment by both actors: the teachers and the learners.

3. Rubrics in L2 writing

In definitory terms, a *rubric* to assess L2 writing is regarded as an assessment tool whose main aim is to evaluate students' writing performance and abilities taking into consideration the type of task and its conventions (Andrade, 2000). Throughout the decades, the use of rubrics have become an essential part of L2 classrooms and L2 language testing as occurs in the case of official language exams such as Cambridge English. As anticipated in the previous section, rubrics are another form of feedback whose scope stretches beyond the mere indication of lexico-morphosyntactic mistakes. In fact, rubrics include detailed feedback and evaluation of the written product as it involves several elements of writing (Sokolik, 2003) which considers the conventions of the genre. Similarly, the use of rubrics does not only involve teacher assessment, but they may also be used as part of self-assessment and peer-assessment. According to Andrade (2006), rubrics contribute to raising more awareness about the quality of students' works, focusing on the different elements of writing rather than on the grade *per se*.

Initially, rubrics in the L2 classroom were transferred from L1 writing instruction as one of the ways in which the evaluation of students' writing could be standardized. L2 teachers have found difficulty in the grading process since, despite having at their disposal assessment tools that are objectively defined, the shadow of subjectivity and bias has traditionally precluded teachers from falling into these issues (Ferris & Hedgcock, 2014). Rubrics provide L2 teachers with a reliable and objective tool in order to examine and further evaluate L2 writing, thus enticing them to adapt pre-existing rubrics (Crusan, 2010) since genre analyses are regarded as extremely time-consuming and out of reach for L2 teachers (Hyland, 2004).

Previous research has stated that rubrics are functionally useful when they include at least three to five evaluative criteria (Popham, 1997). However, more criteria would only preclude teachers from rating properly on the basis of the attributes of each item. In this sense, evaluative criteria within the evaluation of EFL writing have traditionally attempted to focus on content, organization, vocabulary, language use and mechanics (see Jacobs *et al.*, 1981). Although other characterizations may exist, and thus they may adopt different names, the purpose remains the same regarding the evaluative focus.

Rubrics have been extensively used in the L2 classroom as part of the testing system. The scientific literature and language practitioners have traditionally identified two types: (a) holistic rubrics, and (b) analytic rubrics. In the case of holistic rubrics, L2 teachers are expected to assign a *single* score to a specific task. However, Davis and Kondo-Brown (2012) point out that holistic rubrics are not entirely reliable since other aspects of the writing performance are not taken into consideration. As expected, variations exist among the different elements of language: for instance, an L2 writer may not be as skilled at using lexis as in terms of organization. This, again, poses a challenge when using holistic rubrics since the rater has to stick to a single score which may not be a truthful indicator of the performance. In this respect, Weigle (2002, as cited in Rock, 2021) also found that holistic rubrics are not entirely efficient since single scores are correlated with handwriting legibility and text length, which are external characteristics not associated with L2 proficiency.

Analytic rubrics are characterized by a more small-scale evaluation of the criteria since different scores are assigned to a specific competence or element of writing. In essence, analytic rubrics are reported to be more effective in giving more detailed feedback on the various dimensions of writing performance, e.g. text organization, grammar and vocabulary, among others (see Brown, 2012 or Ferris & Hedgcock, 2014). The provision of feedback through an analytic rubrics provides students with a more clear-cut diagnosis of the elements of writing in need of improvement (Rock, 2021). More importantly, the role of analytic rubrics is reflected in the fact that, when L2 writers revise their texts after being provided with feedback, it is much easier to observe the areas in which there is improvement. It is not surprising, though, that analytic rubrics are more representative of a writer's performance. These are usually preferred for placement and diagnostic purposes (Jacobs *et al.*, 1981). Nevertheless, their pedagogical usefulness has been more widespread in the classroom environment owing to the easiness of use and, as mentioned previously, the more detailed feedback they offer. Likewise, previous research has indicated

that rater reliability when using analytic rubrics is substantially high (Knoch, 2009). In a similar vein, Barkaoui (2010) examined how variable rating processes were by taking into consideration the rater experience and the type of rating scale. Results demonstrated that analytic rubrics encouraged raters to articulate and justify scores in a more frequent manner. Such an assessment behavior equally contributed to more internal consistency in ratings.

There have been, however, critical voices suggesting that analytic rubrics are useful as long as their pedagogical value as tools is heuristically analyzed. Turley and Gallagher (2008) proposed four points to examine this value: (1) the purpose of the rubric, (2) the context of use, (3) the role of the rater, and (4) the ideological agenda. These observations are reflected in how dependent the rubric is on the task that L2 learners have to write. In this respect, two distinctions have been made: task-dependent and task-independent rubrics. While the first entails that the scale descriptors are tailored to the specific conventions of the task, i.e. if it is a proposal, the descriptors will include headings as part of the text organization dimension. In the case of task-independent rubrics, the importance is placed on the general L2 proficiency ability, allowing for the generalization of linguistic skills across different tasks. Nevertheless, as pointed out by Rock (2021), task-independent rubrics may fail to display the nuances in genre-sensitive texts. Despite this obvious shortcomings in both type of rubrics, previous research has demonstrated that both task-dependent and task-independent rubrics are equally useful at the time of identifying levels of writing ability (Norris *et al.*, 2002).

Thus, for the present study, one may consider the effectiveness of using task-independent analytic rubrics as the framework of reference. Although rubrics may present some limitations, they still provide learners with the opportunity to express themselves individually to meet the task criteria. In essence, while rubrics appear to be a rather constrained alternative to giving feedback (Hyland, 2004), their validity is reflected in the design of the rubric and the fact that L2 teachers may accompany each of the descriptors with their own feedback comments. Thus, rubrics-based assessment are suggested to contribute to enhancing L2 learners' understanding of the conventions of the task along with the linguistic skills to be improved (Brown, 2012).

4. Learners' use of rubrics in L2 writing

In the previous section, the characteristics of rubrics have been presented along with a series of voices against and for their use. In this section, our focus will be placed on identifying the pedagogical value of using rubrics in the L2 classroom, and more specifically, for the assessment of L2 writing.

Going back to the initial definition of a writing rubric, these are assessment tools whose purpose is to describe and evaluate the students' writing ability in a writing task (Andrade, 2000). Originally transferred from L1 writing instruction, the place of rubrics in the L2 writing classroom has allowed for a standardization of how students are evaluated with regards to their writing ability. Likewise, this standardization paved the way for an "increased sense of transparency" (Ene & Kosobucki, 2016, p. 4), thus helping L2 writers to gain more insight into the changes or perspectives to be taken when engaging in writing. This perception of transparency further enhances the L2 writers' ability to confront internal variables such as anxiety, contributing to its further decrease since the criteria are very clearly laid out.

Added to the previous arguments, rubrics in the L2 classroom provide with a pedagogical value that goes beyond the mere access to the identification of weaker linguistic and content-based areas in writing performance. In this sense, learners' appropriation of the writing conventions of the type of text are also conveyed in the rubric. Thus, more insight is gained into the intricacies of the macrotext beyond superficial linguistic aspects. Caution should be taken, however, in terms of the use of rubrics by both L2 teachers, and also L2 learners. In light of this, rubrics might also refrain L2 writers from focusing on a variety of aspects if these are not specifically included among the criteria or items. This may be reflected in the manner in which L2 writers may hyper-focus on the aspects highlighted in the rubric, leaving aside other important elements of the type of writing (Kohn, 2005; Wilson, 2006).

The affordances offered by rubrics in the L2 classroom are manifold, some of which are: self-assessment, peer-assessment, and teacher assessment (Andrade, 2006). By introducing the value of rubrics beyond the mere assessment or grading more emphasis is placed on the writing conventions, the audience, and the improvement of the learners' writing performance in the L2. In essence, techniques such as peer-assessment or self-assessment, or even assessment exercises with sample writing tasks, allow L2 learners to consider the approach to writing in a different manner.

Within this context, studies focusing on peer feedback have been manifold in L2 writing, in which the learners' feedback was compared with the teacher's, observing whether judgments were similar in terms of criteria (see Topping, 2003). In fact, peer-assessment and self-assessment have received numerous scholarly support from theories such as socio-cultural and activity theories (Swain, 2006). In this context, collaboration plays a pivotal role in advancing language learning since it constitutes a facilitator for scaffolding learning, especially in peer-assessment (Schunk, 2014).

As regards self-assessment, it is an alternative assessment method in L2 writing instruction since it allows learners to become involved in their own evaluation. Prior to the use of self-assessment as an ecologically valid technique, Fahimi and Rahimi (2015) indicated that students were specifically instructed on how to assess their writings by using predefined criteria as shown in the rubrics. Such an assessment task was positively regarded by students. This positive attitude was equally observed in Andrade *et al.* (2009), who explored how the use of rubrics on the long and short term contribute to increasing L2 writers' self-efficacy. Nevertheless, Covill (2012) explored in his study how the absence or use of rubrics and self-assessment were distinctive with regards to L1 writing and L2 writing in Higher Education. In the case of the latter, L2 writers still require additional training on how to use criteria-references rubrics.

Despite the vast amount of research that has examined the value of rubrics as instructional and evaluation tools in the L2 writing classroom, none of it has explored the manner in which learners get accustomed to using rubrics. Such an instruction on the use of rubrics contributes to expanding L2 learners' knowledge about what they are expected to achieve in their writing tasks and the criteria to be met. On the basis of this reasoning, the present study aims to fill this gap by presenting an instructional approach to using writing tasks. Through assessment exercises, L2 writers gained a more in-depth insight about the expectations in rubrics and the connection between criteria and writing performance.

5. Aims and research questions

This study intends to explore how accurate undergraduate students' perceptions may be when evaluating two sample writing tasks with different writing performance.

The research questions that guided this study are the following ones:

1. What are the students' assessments of each sample article (A and B)?
2. To what extent are there any differences in the students' assessment between Article A and Article B?
3. Is there a relationship between the assessment criteria in the rubric in Article A? And in Article B?

6. Method

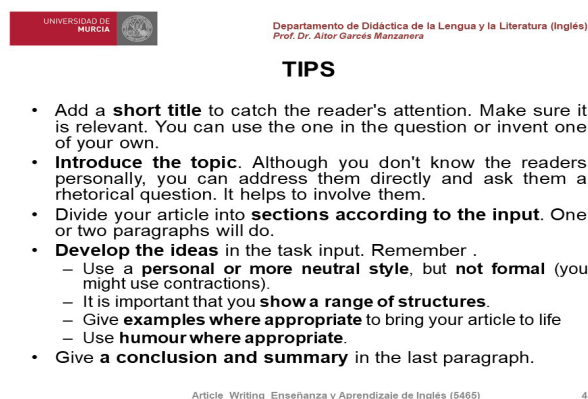
6.1. Context and participants

The participants in our study were part of a total cohort of 41 undergraduate students (male= 8; female= 33). They were attending an EFL module in their third year of undergraduate studies in Primary Education at a public university. This EFL module included contents related to a B1+ level of English according to the Common European Framework of Reference for Languages (CEFRL). They received EFL lessons two hours and a half. Out of these, one hour and a half were devoted to theoretical aspects of the language. The remaining hour was used for speaking practice.

6.2. Data collection procedure

The research design of this study entailed a dual-event data collection procedure. Throughout a one-hour L2 writing instruction lesson, students were taught the writing conventions of an *article* following the guidelines proposed by the well-known testing options such as Cambridge English. To do so, a presentation was shown in class (see Figure 1) in which specific aspects concerning articles as a type of writing were introduced and carefully explained.

Figure 1. Sample slide of the presentation



The slide features a red header with the University of Murcia logo and the text 'UNIVERSIDAD DE MURCIA'. To the right, it identifies the 'Departamento de Didáctica de la Lengua y la Literatura (Inglés)' and 'Prof. Dr. Aitor García Manzanares'. The main content is titled 'TIPS' and lists seven bullet points: 1. Add a short title to catch the reader's attention. 2. Introduce the topic. 3. Divide the article into sections according to the input. 4. Develop the ideas in the task input, with sub-points on style, structures, and humor. 5. Give a conclusion and summary in the last paragraph. At the bottom, it includes the text 'Article_Writing_Enseñanza y Aprendizaje de Inglés (5465)' and the number '4'.

Source: own elaboration

This presentation included essential expressions used for subheadings as well as useful expressions in order to engage the reader, which is part of the textual functions of articles as a communicative text. Once all the guidelines had been thoroughly outlined, students were encouraged to scan a QR code in order to do the activity. This QR code led them to a form in which they had to assess a sample answer to an article task (see Figure 2 below):

Figure 2. Slides showing the task instructions and sample answer A

SAMPLE ARTICLE A INSTRUCTIONS

- You see this announcement in an international magazine.

Articles wanted
Life on a desert island
 Imagine you were on a desert island. What important object, person or place in your life would you miss most? What would be the reasons? Write us an article answering these questions. We will publish the best articles in the magazine.

- Write your **article** in **140-190 words** in an appropriate style.

ARTICLE A

Life on a desert island
 How would you feel about living on a desert island? I can't imagine anything worse! I'd miss a lot of things but most of all, I would miss my home.
 My home is a small house on the outskirts of a city. It was built about 50 years ago and has a small garden. In the summer, our country gets very hot but our house is always cool. You'd probably think our house is nothing special, but I have lived there all my life and all my friends live nearby. It's a happy place, where I feel completely safe. Whenever I go away, I look forward to coming back, lying on my bed, reading a book and listening to my brother and sister arguing downstairs!
 I love travelling and meeting new people, but if I were on a desert island, I'd be away from the place I love most: my home; and I would hate that.

Source: own elaboration

Following a two-stage procedure, students first assessed sample article A using the rubric that they accessed through the QR code. This sample article A was conceived as an ideal response to the writing task, and as such, it was meant to foster warming up their perception of the assessment. Afterward, sample article B was presented, as shown in Figure 3 below:

Figure 3. Slides showing the task instructions and sample answer B

SAMPLE ARTICLE B INSTRUCTIONS

4. You see this notice in your school magazine.

Articles wanted
How to have fun without spending too much money
Life can be expensive – especially for teenagers!
 Tell us how you and your friends have fun without spending a lot of money. The best article will be published in next month's magazine.

Write your **article**.

ARTICLE B

By watch a nature and birds are the best to take pleasure there is different feeling to have fun in nature. Nature make us happy and relive stressful. I think to Have fun it would be better to stay with nature because al it is the best place to have fun.
 So, It would be better to have fun in Nature with friends.
 W We went this last Saturday to Chiseni for traking and which we was climbing that we saw many trees and fruits we also saw water fall and river we play there and by on the grass on the top mid top of at hill there was picnic spot where many people was doing picnic there. ~~Wheld~~ we also joint them as they invited to join them we had a lot of fun there by dancing sing and By having delicious food so, ~~I think-I~~ suggest you to go such place where we can have a lot of fun and less money can be spend.

Sample article B appears to be less cared for since there are numerous expressions and words which are crossed off. Similarly, accuracy and text organization is not the expected one. Students were thus told to evaluate this sample article using the same rubric they had resorted to in the previous sample article A.

6.3. Data collection instruments

In the previous section, two data collection instruments were mentioned, that is, the *PowerPoint presentation* in which the conventions of writing an article at a B2 level were introduced, and the *QR codes*, to access the rubric. In this subsection, a more detailed outline of the rubric will be presented. As can be observed in Figure 4 below, the scale of the rubric is situated between 0 and 5, and it contains five different assesment criteria. Each of these scales provides a series of characteristics to assess the writing task. The first row corresponds to *Grammar*, the second one to *Vocabulary*, and the third row to *Content*. The fourth row aims at evaluating *Communicative Achievement*, and finally, the fifth row is related to *Organization*. Although not entirely, much of this rubric was based on the assessment criteria included in the official Cambridge English: First (B2) examination.

Figure 4. Dimensions with each assessment criterion.

0	1	2	3	4	5
Calificación por debajo de '1'	Usa formas gramaticales simples con buen control. Aunque los errores son claramente visibles, el significado de la oración puede percibirse.	<i>Características compartidas entre '1 y '3'.</i>	Se utiliza una variedad de formas gramaticales simples y complejas con un buen control. Los errores presentes no impiden la comunicación.	<i>Características compartidas entre '3 y '5'.</i>	Se usa un amplio rango de formas simples y complejas con control y flexibilidad. Puede haber errores ocasionales, pero no impiden la comunicación.
Calificación por debajo de '1'	Se usa vocabulario diario de forma generalmente apropiada, aunque a veces se repitan muchas palabras.	<i>Características compartidas entre '1 y '3'.</i>	Se usa una variedad de vocabulario de forma apropiada, con uso ocasional de léxico menos común de forma inapropiada.	<i>Características compartidas entre '3 y '5'.</i>	Se usa una variedad de vocabulario, incluido léxico menos común, de forma apropiada.
El contenido es irrelevante. El lector clave no es informado.	Hay contenido irrelevante y la tarea no se ha interpretado bien. El lector clave es informado mínimamente.	<i>Características compartidas entre '1 y '3'.</i>	Hay omisiones de carácter menor o de poca relevancia. El lector es informado por completo.	<i>Características compartidas entre '3 y '5'.</i>	La intención es clara y todo el contenido es relevante para la tarea. El lector comprende a la perfección la misma y es informado por completo.
Calificación por debajo de '1'	Usa las convenciones de la tarea de forma generalmente adecuada para comunicar las ideas claramente.	<i>Características compartidas entre '1 y '3'.</i>	Usa las convenciones de la tarea para mantener la atención del lector y comunicar ideas de forma clara.	<i>Características compartidas entre '3 y '5'.</i>	Usa las convenciones de la tarea de forma efectiva para mantener la atención del lector y comunicar ideas complejas y directas de forma apropiada.
Calificación por debajo de '1'	El texto está conectado de forma coherente a través de conectores básicos y un número limitado de marcadores cohesivos.	<i>Características compartidas entre '1 y '3'.</i>	El texto está, por lo general, bien conectado y organizado, usando diversos conectores y marcadores cohesivos.	<i>Características compartidas entre '3 y '5'.</i>	El texto está muy bien organizado y coherentemente conectado a través de diversos marcadores cohesivos y patrones organizacionales, que crean un buen efecto en el lector.

Source: own elaboration

The rubric was integrated into a *Google Forms* questionnaire in order to facilitate this task. Before proceeding with the selection and evaluation of each dimension, students had to choose the article they were assessing (i.e. Article A or B).

6.4. Statistical analyses

The statistical analyses proposed in the study were performed according to the research questions presented in section 5. Thus, in terms of the first research question, descriptive statistics were used in order to present the data. The mean, the median, the standard deviations and minimum and maximum were used to describe the raw data from the assessment of each article.

As regards the second research question, an independent samples Welch's test was performed in an attempt to observe whether there were differences between the assessment ratings in Article A and B. Since the data had not been paired – the assessment had been kept anonymous – we opted for independent samples tests. Welch's T test was used since the variances of the data were not equal, and the distributions were not normal for all the dimensions. In order to observe the magnitude of the effect, Cohen's d was calculated. Following Plonsky and Oswald (2014), that is, their benchmarks for effect sizes in L2 research, the results were calculated on the basis of these benchmarks: small (.40), medium (.70), and large (1.0). Finally, as for the third research question, a series of Spearman's Rho correlations were performed on the data individually for the assessment of Article A and B. Spearman's Rho correlations were used since our variables of interest were continuous with outliers in addition to the small sample size of our study. All these analyses were performed using the freely accessible software packages JASP 0.16.3. and Jamovi 2.3.13.

7. Results

RQ1. What are the students' assessments of each sample article (A and B)?

As can be observed in Table 1, undergraduate students coincided in assessing with higher marks Article A. Although in all the dimensions the differences are considerable, it does not seem to be the case in terms of *Content*, whose values are very similar. Along this lines, it must be noted that *Communicative Achievement* was assessed with an extremely low mark in Article B. It seems to be the case of *Organization* as well if compared with Article A.

Table 1. Descriptive statistics for each dimension of the rubric and the writing task evaluated.

		N	Mean	Median	SD	Minimum	Maximum
Content	Article A	41	2.220	2.00	1.388	0.00	5.00
	Article B	39	1.692	1.00	1.260	0.00	4.00
Communicative Achievement	Article A	41	2.317	2.00	1.059	0.00	4.00
	Article B	39	0.744	1.00	0.637	0.00	2.00
Organization	Article A	41	2.244	2.00	1.338	0.00	5.00
	Article B	39	0.308	0.00	0.569	0.00	2.00
Grammar	Article A	41	3.024	3.00	1.151	1.00	5.00
	Article B	39	0.949	1.00	0.560	0.00	2.00
Vocabulary	Article A	41	2.610	2.00	0.972	1.00	5.00
	Article B	39	1.359	1.00	0.811	0.00	3.00

RQ2. To what extent are there any differences in the students' assessment between Article A and Article B?

Table 2 displays the results corresponding to the Welch's t inferential test. Let us commence with *Content*. As it has been anticipated in the previous research question, this dimension did not appear to be as dissimilar as the other ones. The mean difference is not very high (MD= 0.527) and the effect size is trivial (d = 0.398). Moving on to *Communicative Achievement*, the test yielded a significant result (p < .001), with a considerable mean difference (MD= 1.573). This is also reflected in the large effect size (d = 1.80).

In the case of *Organization*, the mean difference is even more marked (MD = 1.936), and the result is statistically significant (p < .001). The effect size is large (d = 1.884). Such a difference between the assessment of Article A and B is much more considerable in the case of *Grammar*, whose result is statistically significant (p < .001), and the mean difference is higher than in the rest (MD = 2.076). The effect size is very large (d = 2.294), which indicates that the means are different in nature. Finally, in terms of *Vocabulary*, the result is statistically significant (p < .001) with a moderate mean difference (MD = 1.251). This is equally reflected in the large effect size (d = 1.398).

Table 2. Welch's t tests results for each dimension.

		Statistic	df	p	Mean difference	SE difference	Effect Size
Content	Welch's t	1.78	77.8	0.079	0.527	0.296	Cohen's d 0.398
Communicative Achievement	Welch's t	8.10	66.2	<.001	1.573	0.194	Cohen's d 1.800
Organization	Welch's t	8.50	54.6	<.001	1.936	0.228	Cohen's d 1.884
Grammar	Welch's t	10.34	58.5	<.001	2.076	0.201	Cohen's d 2.294
Vocabulary	Welch's t	6.26	76.7	<.001	1.251	0.200	Cohen's d 1.398

RQ3. Is there a relationship between the assessment criteria in the rubric in Article A? And in Article B?

In this research question, we aimed to provide an answer to the internal relationships existing among the different dimensions. Table 3 displays the correlations for Article A. As can be observed, *Content* is positively correlated with *Communicative Achievement* and *Organization* (rs = .570 and .727, respectively), being these statistically significant. In turn, *Communicative Achievement* is strongly correlated with *Organization* (rs = .629), while in the case of *Grammar* and *Vocabulary*, correlations are moderately and weakly correlated (rs = .486 and .385, respectively). Interestingly, *Grammar* is moderately correlated with both *Organization* and *Vocabulary* (rs = .442 and .476, respectively).

Table 3. Spearman’s correlations for Article A.

Spearman’s Correlations ARTICLE A			Spearman’s rho	p
Content	-	Communicative Achievement	0.570	*** < .001
Content	-	Organization	0.727	*** < .001
Content	-	Grammar	0.494	** 0.001
Content	-	Vocabulary	0.362	* 0.020
Communicative Achievement	-	Organization	0.629	*** < .001
Communicative Achievement	-	Grammar	0.486	** 0.001
Communicative Achievement	-	Vocabulary	0.385	* 0.013
Organization	-	Grammar	0.442	** 0.004
Organization	-	Vocabulary	0.267	0.092
Grammar	-	Vocabulary	0.476	** 0.002

* p < .05, ** p < .01, *** p < .001

Moving on to the Spearman’s correlations for Article B, the results are presented in Table 4 below. As observed, correlations are overall very weak, indicating that the degree of interrelation among the dimensions is dissimilar in the case of Article B. The only exception is observed between *Communicative Achievement* and *Organization*, whose correlation is weak (rs = .350).

Table 4. Spearman’s correlations for Article B.

Spearman’s Correlations ARTICLE B			Spearman’s rho	p
Content	-	Communicative Achievement	-0.026	0.875
Content	-	Organization	0.037	0.822
Content	-	Grammar	0.018	0.911
Content	-	Vocabulary	0.008	0.961
Communicative Achievement	-	Organization	0.350	* 0.029
Communicative Achievement	-	Grammar	0.228	0.163
Communicative Achievement	-	Vocabulary	0.198	0.226
Organization	-	Grammar	0.201	0.221
Organization	-	Vocabulary	0.265	0.102
Grammar	-	Vocabulary	0.236	0.148

* p < .05, ** p < .01, *** p < .001

8. Discussion

The present study found that, despite common beliefs in the misuse of rubrics in the EFL classroom tied with the increased feeling of anxiety (Arindra & Ardi, 2020), undergraduate students were skilled – to some degree – when it came to assessing sample writings as part of their writing practice. In our study, Article A was the ideal response to the task instructions while Article B clearly included aspects of language in need for improvement. An important finding in this case was related to the dimension of *Content*, which the undergraduate students were able to assess correctly (Article A= 2.22; Article B= 1.69), not letting their assessments be clouded by the other dimensions. Previous research, such as Reddy and Andrade’s (2010), indicated that the use of rubrics contributed to increasing transparency as well as the quality of the written product, guiding not only peer assessment but also self-assessment. Equally important, the participants were able to identify aspects of *Vocabulary* (Article A= 2.61; Article B= 1.35), with important differences between both articles. The fact that they were presented with an

ideal response to the writing instructions in Article A did not contribute to contaminating their views at the time of assessing Article B. In an area such as vocabulary, where boundaries may be blurred in assessment, especially for undergraduate students whose expertise in assessing pieces of writing is not high, this demonstrates that defining the aims of rubrics is certainly beneficial.

Additionally, the participants were able to distinguish the intricacies in the dimension of *Grammar* with a degree of accuracy which would not be expected by B2 level students. In that case, their observations indicated a high mark in the case of Article A (M= 3.02) while it was far lower in Article B (M= 0.94). It must be noted that Article B had been obtained from the *Cambridge B2 First Handbook* (available at their website), in which they already offered the assessment of the piece of writing by dimensions. Surprisingly, undergraduate students were more demanding when evaluating this aspect of language in comparison with Cambridge's own assessment, which was set at Band 2. Nonetheless, such a finding equally provides us with an interesting perspective: firstly, undergraduate students seem to approach assessment in a harder and more demanding manner given the low marks, and second, it also hints at the students' ability to correctly stick to the scale. Another potential explanation behind this finding may be rooted in the teacher's expectations of their grading (Whittaker *et al.*, 2001) since participants might have been much harder upon correcting in an attempt to cover their weaknesses at identifying linguistic and content issues.

As mentioned previously, the greater differences were observed in Grammar and Organization. This clearly reflects the students' awareness about text organization, being it one of the most visual features when assessing a piece of writing.

Moving on to the internal relationship of the assessment in relation to writing conventions, we found out that *Content* was highly correlated with *Organization* in the case of Article A, that is, the sample writing answer which was tailored to achieve the maximum mark in all the dimensions of the rubric. Similarly, *Communicative Achievement* was found to bear a strong relationship with *Organization*. These diverse results lead us to consider that L2 learners, when using rubrics, rely too much on the visual appearance of the text. This visual reliance is weakly associated with linguistic aspects such as *Grammar* or *Vocabulary*. Again, the positive observation of the rest of the dimensions may have had an influence on how students might have assessed the other criteria.

It must be noted, though, that these associations were not replicated in the case of the assessment of Article B, that is, the article that deliberately included varied issues. The only finding to be highlighted is a weak association between *Communicative Achievement* and *Organization*, which leads us to insist on the influence exerted by the visual appearance and the initial screening of the text.

9. Conclusion

This study has been an attempt to explore undergraduate students' ability to comprehend the intricacies of rubrics and use them as part of an exercise. Additionally, our study has allowed us to gain rich insight into the manner in which tertiary level students may understand the assessment process through the use of rubrics. Our first objective was to provide a descriptive overview of the assessment of two sample writing tasks, and the second main objective was to observe whether there were significant differences among them, and ultimately, to detect any type of association therewith.

Several conclusions can be drawn: on the one hand, undergraduate L2 students are able to identify differences at micro- and macro-level, especially when two writing tasks are distinct in terms of quality. Second, assessing the writing tasks has demonstrated that L2 students are, to some extent, unable to comprehend some of the categories or at least partially since their assessment is more demanding and harder. This, in some sense, may impair the use of rubrics as part of self-assessment as meaningful resources (Brooks, 2012). Thus, writing instruction should care for a pre-training in which the criteria of the rubric is made clear and explicit in order to ensure that L2 students are able to use the information provided in an efficient manner. Likewise, the visual appearance and the initial impression of the writing task plays an important role in determining how other criteria are assessed. Such a finding may point to the necessity of underscoring the individualism of each criterion. In other words, the fact that *Content* may be fully covered does not entail that *Communicative Achievement* involves a positive grade.

These are tentative conclusions since this study was mostly based on their performance following an exercises which involved the assessment of two writing tasks with the same writing conventions. Nevertheless, it has allowed us to observe that L2 students at a tertiary level are more than capable of identifying the quality of a writing task by means of a rubric.

Our study is not without its limitations. Firstly, our pool of participants was particularly limited (N= 41), making it a necessary to expand the number of participants in order to confirm solid generalizations. Additionally, more L2 writing tasks should be included in the research equation since L2 students may be more skilled at assessing some tasks over others. Similarly, the inclusion of more L2 proficiency levels should be a major concern in future research avenues, as this could give us rich data about the extent to which proficiency may be an influential variable.

List of references

- Andrade, H. G. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57(5), 13–19.
- Andrade, H. L. (2006). The trouble with a narrow view of rubrics. *The English Journal*, 95(6), 9. <https://doi.org/10.2307/30046616>
- Andrade, H. L., Wang, X., Du, Y., & Akawi, R. L. (2009). Rubric-referenced self-assessment and self-efficacy for writing. *The Journal of Educational Research*, 102(4), 287–302. <https://doi.org/10.3200/JOER.102.4.287-302>
- Arindra, M. Y., & Ardi, P. (2020). The correlation between students' writing anxiety and the use of writing assessment rubrics. *LEARN Journal: Language Education and Acquisition Research Network*, 13(1), 76–93.
- Barkaoui, K. (2010). Variability in ESL essay rating processes: The role of the rating scale and rater experience. *Language Assessment Quarterly*, 7, 54–74. <https://doi.org/10.1080/15434300903464418>
- Bitchener, J., & Storch, N. (2016). *Written Corrective Feedback for L2 Development (Vol. 96)*. Multilingual Matters.
- Brown, J. D. (2012). Developing rubrics for language assessment. In J. D. Brown (Ed.), *Developing, using, and analyzing rubrics in language assessment with case studies in Asian and Pacific languages* (pp. 1–9). University of Hawaii National Foreign Language Resource Center.
- Byrnes, H. (2012). Conceptualizing FL writing development in collegiate settings: A genre-based systemic functional linguistic approach. In R. M. Manchón (Ed.), *L2 writing development: Multiple perspectives* (pp. 191–219). Walter de Gruyter.
- Covill, A. E. (2012). College students' use of a writing rubric: Effect on quality of writing, self-Efficacy, and writing practices. *The Journal of Writing Assessment*, 5(1), 1–9.
- Crusan, D. (2010). *Assessment in the second language writing classroom*. University of Michigan Press.
- Davis, L., & Kondo-Brown, K. (2012). Assessing student language performance: Types and uses of rubrics. In J. D. Brown (Ed.), *Developing, using, and analyzing rubrics in language assessment with case studies in Asian and Pacific languages* (pp. 33–55). University of Hawaii National Foreign Language Resource Center.
- 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. [doi:10.1016/j.asw.2016.06.003](https://doi.org/10.1016/j.asw.2016.06.003)
- Fahimi, Z. & Rahimi, A. (2015). On the Impact of Self-assessment. Practice on Writing Skill. *Social and Behavioral sciences*, 192, 730–736
- Ferris, D. R. (2014). Responding to student writing: Teachers' philosophies and practices. *Assessing Writing*, 19, 6–23. <https://doi.org/10.1016/j.asw.2013.09.004>
- Ferris, D. R., & Hedgcock, J. S. (2014). *Teaching L2 composition: Purpose, process, and practice*. Routledge.
- Hyland, K. (2004). *Genre and second language writing*. University of Michigan Press
- Hyland, K. (2016). Methods and methodologies in second language writing research. *System*, 59, 116–125. <https://doi.org/10.1016/j.system.2016.05.002>
- Jacobs, H., Zinkgraf, S., Wormuth, D., Hartfiel, V., & Hughey, J. (1981). *Testing ESL composition: A practical approach*. Newbury House.
- Knoch, U. (2011). Rating scales for diagnostic assessment of writing: What should they look like and where should the criteria come from? *Assessing Writing*, 16, 81–96. <https://doi.org/10.1016/j.asw.2011.02.003>
- Kohn, A. (2006). The trouble with rubrics. *English Journal*, 95(4), 12–15.
- Leki, I., Cumming, A., & Silva, T. (2008). *A synthesis of research on second language writing in English*. Routledge.
- Leow, R. P. (2020). L2 writing-to-learn. In R. M. Manchón (Ed.), *Writing and Language Learning: Advancing research agendas* (pp. 95–118). Amsterdam: Benjamins.
- Manchón, R. M. (2009). Broadening the perspective of L2 writing scholarship: The contribution of research on foreign language writing. In R. M. Manchón (Ed.), *Writing in foreign language contexts: Learning, teaching, and research* (pp. 1–19). Multilingual Matters.
- Manchón, R. M. (2011). Writing to learn the language: Issues in theory and research. In R. M. Manchón (Ed.), *Learning to write and writing to learn in an additional language* (pp. 61 – 82). Amsterdam: Benjamins.
- Manchón, R. M., & Vasylets, O. (2019). Language learning through writing: Theoretical perspectives and empirical evidence. In J. B. Schwieter & A. Benati (Eds.), *The Cambridge handbook of language learning* (pp. 341–362). Cambridge: CUP. <https://doi.org/10.1017/9781108333603.015>
- Manchón, R.M. & Polio, C. (Eds.). (2021). *The Routledge Handbook of Second Language Acquisition and Writing*. Routledge
- Martin, J. R. (1999). Modelling context: A crooked path of progress in contextual linguistics. In M. Ghadessy (Ed.), *Text and context in functional linguistics* (pp. 25–61). John Benjamins.
- Martin, J. R. (2002). Meaning beyond the clause: SFL perspectives. *Annual Review of Applied Linguistics*, 22, 52–74. <https://doi.org/10.1017/S026719050200003X>
- Norris, J. M., Brown, J. D., Hudson, T. D., & Yoshioka, J. (2002). *Designing second language performance assessments*.

University of Hawaii.

- Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language learning*, 64(4), 878-912.
- Polio, C. (2012). The relevance of second language acquisition theory to written error correction debate. *Journal of Second Language Writing*, 21, 375-389.
- Popham, W. J. (1997). What's wrong-and what's right-with rubrics. *Educational Leadership*, 55, 72-75. <https://www.ascd.org/el/articles/whats-wrong-and-whats-right-with-rubrics>
- Reddy, Y.M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435-448.
- Rock, K. (2021). *Using Analytic Rubrics to Support Second Language Writing Development in Online Tasks*. [Doctoral Thesis]. Available online. University of Hawai'i at Manoa
- Schunk, D. H. (2014). *Learning theories: An educational perspective*. Pearson College Division.
- Sokolik, M. (2003). Writing. In D. Nunan (Ed.), *Practical English language teaching* (pp. 87-108). New York: McGraw-Hill
- Swain, M. (2006). Linguaging, agency and collaboration in advanced language proficiency. In H. Byrnes (Ed.), *Advanced language learning: The contribution of Halliday and Vygotsky* (pp. 95-108). Continuum.
- Tardy, C. M. (2012). A rhetorical genre theory perspective on L2 writing development. In R. M. Manchón (Ed.), *L2 writing development: Multiple perspectives* (pp. 165-190). Walter de Gruyter.
- Topping, K. (2003). Self- and peer assessment in school and university: Reliability, validity and utility. In M. Segers, F. Dochy, & E. Cacallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards* (pp. 55-87). Dordrecht: Kluwer Academic Publishers.
- Turley, E., & Gallagher, C. (2008). On the uses of rubrics: Reframing the great rubric debate. *English Journal*, 97(4), 87-92.
- Wilson, M. (2007). Why I Won't Be Using Rubrics to Respond to Students' Writing. *The English Journal*, 96(4), 62-66.