

## Use Of Jacobs ESL Composition Profile to Evaluate University Students' Writing

**Hari Prasad Tiwari**

Email: [haritiwarimmc@gmail.com](mailto:haritiwarimmc@gmail.com)

Lecturer of English Education Tribhuvan University, Mahendra Multiple Campus,  
Nepalgunj, Nepal

Orcid: 0000-0002-0023-3360

### Abstract

*A teacher should have an access to a reliable scoring rubric in order to assess the ESL/ EFL (English as a Second Language/ English as a Foreign Language) students' writing abilities accurately. One of the most widely used tools for assessing writing skills of non-native speakers of English is Jacobs ESL Composition Profile which can help teachers tailor their instruction to better meet the needs of their students. The present study attempts to determine the internal consistency between teachers i.e. raters who use Jacob's Scoring Rubric (SR) to evaluate students' essays, describe the level of students' writing performance as measured by the Jacob's SR, and describe teachers' viewpoints while using the profile. The research utilizes descriptive quantitative research design. The participants consisted of two ESL/EFL teachers who have been teaching and testing writing at bachelor's level for more than five years and 40 fourth year ESL/ EFL students studying at undergraduate level at Madhyaschim Multiple Campus, Nepalgunj Banke, Nepal. The researchers used judgmental sampling to select the participants of the study. The teachers were instructed to score 40 essays written by fourth-year students. The finding revealed High internal consistency ( $r = 0.86$ ,  $\alpha = 0.00 < 0.05$ ) between TA and TB when assessing student essays using the Jacobs ESL Composition Profile. In addition, the Cronbach alpha analysis reveals a value of 0.918, which suggests a high degree of consistency. The writing performance of students was found to be in four categories: very good (3.75 %), good (52.5 %), satisfactory (36.5 %) and acceptable (7.5 %). The outcome of the research demonstrates that the Jacobs ESL Composition Profile remains reliable for essay scoring, even though proficiency and experience are necessary due to its comprehensive guidelines.*

**Keywords:** Essay writing, internal consistency, rater, scoring rubric, opinions

### INTRODUCTION

The Jacobs ESL Composition Profile is an assessment tool designed to evaluate non-native English speakers' writing skills. It was developed by Glenn F. Jacobs, a teacher of TESOL (Teaching English to Speakers of Other Languages), and was first published in 1981. The Jacobs ESL Composition Profile consists of a set of criteria for evaluating different aspects of writing, including content, organization, vocabulary, grammar, and mechanics. It is intended to provide a comprehensive evaluation of a student's writing abilities and to help teachers identify areas where students may need additional instruction and practice.

The profile is typically used by teachers of ESL to assess the writing skills of their students. It can also be used by researchers to study the writing abilities of non-native speakers of English. Overall, the Jacobs ESL Composition Profile is a useful tool for assessing the writing skills of non-native speakers of English and can help teachers tailor their instruction to better meet the needs of their students.

As stated by Bachman and Palmer (1996), assessing students' writing is considered challenging in the context of EFL/ ESL. It requires a lot of effort, time and judgmental ability to evaluate written students' work. EFL/ ESL teachers are often preoccupied with assessing students' grammar, punctuation, spelling, diction, cohesive devices, transitions between paragraphs, important topics, and sentence connectors on a regular basis. Fulcher (2003) asserts that teachers often struggle not just with teaching writing but also with scoring of the students' written work. Teachers must be clear in how to score and how to analyze the writing to avoid the wrong judgments about students' writing abilities,

Given this context, according to Hughes (1989) SR helps teachers consistently assign grades to different sections of student work by outlining specific criteria against which evaluations may be made. Assessing performance is the main goal of the rubric. Knoch, (2011) describes the two main parts of a rubric, which are the criteria themselves and the descriptions of how well they were met. According to Nemati (2007) rubrics allow teachers and instructors to make observations and link them with descriptions, preventing them from making hasty judgements in class

### Literature Review

As stated by Alderson et al. (1995) Primary Trait SR (PTSR), Holistic SR (HCR), and the Analytic SR (ACR) are three types of SRs which can be used by teachers to assess students' writing proficiency in a SL. PTSR places an emphasis on particular rhetorical talents in writing (Weigle, 2002). It is clear and easy to use since it just requires the instructor to offer a single score to the portion of the text that is deemed to be the most relevant. However, according to Zomorodian (1998), it is not sensitive enough to differentiate between compositions that are well-developed and those that are under-developed, and thus may not be appropriate for evaluating the final drafts of students' work. Regarding this topic, Weir (1990) said that PTSR system is most beneficial when it is used for delivering comments on student drafts as well as inspiring and shaping revisions.

The second kind of SR is known as HCR. It evaluates the overall quality of the composition by applying all of the criteria simultaneously (Shohamy, 1993). This allows the composition of students to be defined by the holistic SR. According to Shaw and Weir (2007), HCR requires teachers to evaluate the work as a whole and provide a single score to reflect the whole of the text being evaluated. Because it assigns a single score to the whole of the work. It is best suited for summative assessment (Knoch, 2007). Despite the fact that it may be challenging to develop and it is less complicated and takes less time. According to Stemler (2004), providing a text with a single score does not provide any advice on how to improve the writing of students' and makes it inappropriate for use in formative assessment. Knoch (2009) further writes that if writing is assessed as a whole, it may not adequately represent the quality of the composition being graded. It is unable to provide much help to the development of the students, particularly for those students whose performance is below average or just average.

ASR is the third kind of grading rubric that may be used. According to Knoch, (2009), ASR is most useful when used in a classroom setting since it tackles each criterion separately. The criteria that are included in the evaluation rubric for analytical writing need to be aligned with the characteristics of writing.

Haswell (2005) suggests that in order to produce a piece of written work, writers need to ensure that their work satisfies six essential requirements of writing. There are six primary components, and they are as follows: organization content, discourse, vocabulary, syntax and mechanics. The content discusses the thesis statement, as well as linked concepts, suggestions for development, and descriptive language. The effectiveness of the introduction, the sequence in which ideas are presented, and the overall length of the document are all aspects of organization. Discourse is the third component, and it includes elements such as the subject

phrase, the unity of the paragraph, transitions, the author of the discourse, coherence, theatrical conventions, references, fluency, economy, and diversity. The fourth component is called "mechanics," and it refers to the use of proper presentation, as well as good spelling, punctuation, and citation of references.

According to Maftoon and Akef (2010), writing ability should not be taught as a series of separate components but rather as an integrated totality. When a student composes a piece of writing, Bachman and Palmer (1996) assert that the learner is actively engaging with all aspects of language at that point in time. In addition, he says that there is a significant probability that writing will be taught in a holistic manner rather than via the mastery of specific talents in the near future. Reid (1993) observes that all throughout the imaginative process of writing, there is an active relationship between the subject matter and the language used. As stated by Norton (2000), the language components (content, organization, vocabulary, language use, and mechanics) interact with one another throughout the process of writing.

A SR that takes into consideration the writing components is called the Jacobs ESL Composition Profile. Jacobs et al. (1981) developed this analytical type profile in the year 1981. Essays are scored based on the following five unique rating categories of writing quality that are included in their rubrics: subject (35 points), organization (20 points), vocabulary (20 points), language use (25 points), and mechanics (5 points). Each of these areas is assigned a different point value. Each group of criteria uses an interval scoring system to translate a subjective rating scale that has four levels. The properties of this profile, which was established by writing researchers employed by a testing firm, make it one of the most extensively used and trustworthy ones for assessing ESL content (Lee et al., 2008). Its characteristics were developed by writing researchers. In addition to this, it is regarded as one of the most well-known rubrics in the field of writing in a second language (Brooks, 2012). According to Hamp-Lyons (1991), the Jacobs ESL Composing profile has been around since 1981, although it has only recently acquired popularity among teachers of second languages and academics in the field.

The assessment of writing incorporates a measure of research, with specific emphasis on the Jacobs ESL composition profile. Bacha (2001) carried out the study, focusing on examining the effectiveness of employing both holistic and analytical evaluation methods for grading EFL students at Lebanese American University. The student cohort comprised entirely of Arab individuals utilizing English as a non-native language. The study employed Jacob's ESL Composition Profile as the analytical scoring framework. The findings of this research indicate that opting for the analytical scoring rubric is a more advantageous decision within the EFL curriculum, as it offers a more comprehensive insight into students' essay proficiency. Klimova (2011) undertakes a comparison between the application of holistic and analytical scoring methods, as formulated by Jacobs et al. (1981), within the Czech ESL setting. The researcher's deduction is that analytical scoring surpasses holistic scoring due to its capacity to provide a richer understanding of students' learning outcomes. Nevertheless, he advocates for the combined utilization of both scoring approaches when assessing lengthier essays. The incorporation of both methods can offer a more comprehensive overview of students' general advancement and the potential errors that can be mitigated.

Ghanbari et al. (2012) discovered that the foundation of Jacobs ESL Composition Profile is uncertain in terms of the origins of its five trait elements. Their comprehensive investigation exposes that among the contributors involved in devising the Jacobs ESL Composition Profile, a few lacked an ESL background and possessed limited experience in Teaching English as a Second Language (TESOL). Additionally, the participants of the compositions were freshmen at Cornell, Middlebury College, and the University of Pennsylvania. The researchers suspect that none of these writers were second language learners. Consequently, they deduce that a scoring framework originating from contributors

without an ESL background could jeopardize its suitability for evaluating EFL writing competence. They propose that a localized writing scale created by local evaluators might yield more dependable results due to its alignment with the specific context.

However, prior research pertaining to Jacobs ESL Composition Profile has not furnished sufficient details concerning the internal agreement among raters and their individual perceptions when employing it within the EFL environment. Investigating this internal coherence among raters can offer valuable insights to writing instructors and evaluators regarding the credibility of the tool for scoring written work. It sheds light on whether the instrument is genuinely dependable and does not lead to misinterpretations during its application. Furthermore, the study's findings are anticipated to provide guidance on enhancing the internal consistency among raters, thus enabling a more accurate acknowledgment of students' genuine achievements. Hence, the current investigation aims to uncover the internal consistency among raters utilizing the Jacobs SR to score students' essays, depict the level of writing proficiency demonstrated by students when evaluated using Jacobs ESL SR, and outline raters' perspectives and experiences while utilizing the profile.

## METHOD

This study was conducted employing a mixed-method approach since both qualitative and quantitative data were used. Instead than employing just one technique, the mixed-method approach enables researchers to combine two types of data to get a deeper knowledge of the study subject. Two teachers and 40 fourth year ESL/EFL students studying English at undergraduate level in Madhyaschim Multiple Campus, Nepalgunj Banke, Nepal participated in the study. I employed judgmental sampling to select the sample and convenience sampling to select the campus. The raters were the teachers with a minimum of five years of experience of teaching and testing writing at bachelor level. Both teachers were male with master's degrees in English education. The selected teachers were named as TA and TB for their anonymity. The researcher employed test and interview as the instruments to collect the required data. The researchers provided two weeks of instruction to the teachers or raters prior to the study. The instruction mostly focused on teachers' comprehension of the specified rating criteria. Each point of the rating criteria was discussed and accompanied with an illustration. Teachers were also tasked with rating composition samples from the same individual to be reviewed. After two weeks of instruction to the raters, one-day instruction was given to the students. After giving clear instructions to the students, they were asked to write essays. Forty minutes' time was provided to the students to complete the essay. Two weeks' time was provided to the teachers to complete the scoring of students' essays. This amount of time was allotted to prevent the pitfalls of either overestimating or underestimating the genuine ability level of the students being evaluated due to the physical and emotional limitations of the teachers. After the teachers individually rated the writing of each student, the results were compiled into a final score. The researchers divided the students' overall performance into seven categories: outstanding (scores between 90-100), excellent (scores between 80-89), very good (scores between 70-79), good (scores between 60-69), satisfactory (scores between 50-59), acceptable (scores between 40-49) and insufficient (less than 39). The Jacobs ESL Composition Profile scores were subject to a ten-point variance tolerance. This means that if there was a disparity more than ten points, a third teacher was asked to evaluate the identical essay. The scores utilized for calculation are the ratings that are the closest between teachers with a ten-point range.

I used descriptive statistics to portray the quantitative data in the form of graphs and tables. I employed straightforward coding in order to classify the qualitative data. The category-specific descriptors were used to assign a separate point value to each individual piece. This study utilized two methods to assess inter-rater reliability. These methods include calculating

the Pearson Product Moment Correlation between the ratings of the first and second raters, as well as using Coefficient Alpha to measure the internal consistency of the final scores when two raters evaluate each essay. The correlation between the raters' scores and the coefficient alpha were both computed using IBM SPSS version 20.0 for Windows.

## FINDINGS AND DISCUSSION

In this section, the quantitative and qualitative results are presented. Quantitative results are presented by highlighting the numerical patterns, showing relationships, and indicating statistical significance identified through rigorous data analysis. Qualitative findings on the other hand are presented in thematic way.

### Internal Consistency between Teachers

Two teachers independently scored the essays written by the students using the Jacobs ESL Composition Profile (Jacobs et al., 1981) based on five criteria: content, organization, vocabulary, language usage, and mechanics. The descriptive statistics of the scores acquired from teachers are shown in the following table.

**Table 1.** *Descriptive Statistics of Score of the Students*

Teacher	N	Min	Max	Mean	Std. Deviation
T1	40	41	65	57.02	6.057365
T2	40	43	69	60.8	7.278719

Valid N (list wise) 40

Prior to the calculation of descriptive statistics, the raw data indicates that six out of 40 student essays were either over or under rated by both teachers. As the gap between teachers was more than the agreed-upon 10 points, a third teacher was asked to evaluate the writings of six students. For the computational analysis, the closest ratings among teachers were used. According to Bacha (2001), if there is a score difference, a third teacher might be called to assess the identical piece, and the closest score is utilized to derive the average score. The mean scores of TA and TB are 57.02 and 60.8 respectively. Thus, the average score of TB is 3.78 points greater than the average score from TA. In terms of standard deviation, the standard deviation from teacher A (6.057365) is less than the standard deviation from TA (7.278719) and has a standard deviation range of 1.221354. Although, the mean of the teachers varies just 3.78 point range disparity. The standard deviation demonstrates that the scores provided by the TB have greater volatility and spread further away from the mean, unlike TA.

**Table 2.** *Correlations in Score*

		R1	R2
R1	Pearson	1	0.863116**
	Sig. (2 tailed)		.000
	N	40	40
R2	Pearson	0.863116**	1
	Sig. (2 tailed)	.000	
	N	40	40

\*\* correlation is significant at the 0.01 level (2-tailed).

The results show a strong and significant positive correlation between two raters. A correlation coefficient is a number that measures the strength and direction of the relationship between two variables, ranging from -1 to 1. A positive correlation means that the variables tend to move in the same direction, while a negative correlation means that they tend to move

in opposite directions. A correlation coefficient close to 1 or -1 indicates a strong relationship, while a coefficient close to 0 indicates a weak or no relationship. The results show that the correlation coefficient between R1 and itself is 0.863116, which is very high and positive, as expected when a variable is correlated with itself. This means that R1 has a strong positive relationship with itself, meaning that higher values of R1 are associated with higher values of R1. Similarly, the correlation coefficient between R2 and itself is also 0.863116, indicating a strong positive relationship within R2, meaning that higher values of R2 are associated with higher values of R2. The correlation coefficient between R1 and R2 is not directly given in the results, but it can be inferred from the symmetry of the correlation matrix. The symmetry means that the correlation coefficient between two variables is the same regardless of their order. Therefore, the correlation coefficient between R1 and R2 is the same as the correlation coefficient between R2 and R1, which is 0.863116. This means that R1 and R2 have a strong positive relationship with each other, meaning that higher values of R1 are associated with higher values of R2, and vice versa. The results suggest that R1 and R2 are highly correlated and move together in the same direction.

**Table 3.** *Reliability Statistics*

Cronbach alpha	Cronbach's Alpha Based on Standardized Items	N of Items
0.918188	0.0835188	2

The result shows a discrepancy between the Cronbach's alpha values for the original and standardized items. Cronbach's alpha is a measure of internal consistency, which indicates how well a set of items are related to each other and measure the same construct. The higher the alpha value, the more reliable the measurement is. According to Bacha (2001), a Cronbach alpha value more than 0.60 is regarded as dependable and consistent. The result shows that the Cronbach's alpha for the original, unstandardized items is 0.918188, indicating high internal consistency. This means that the original items are very similar and measure the same construct in a consistent way. However, when the items are standardized, Cronbach's alpha drops significantly to 0.0835188, suggesting poor internal consistency among the standardized items. This means that the standardized items are very different and measure different constructs in an inconsistent way. Standardizing the items means transforming them to have a mean of zero and a standard deviation of one, which makes them more comparable. However, standardizing the items may also change their relationships and meanings, resulting in lower internal consistency. Therefore, the result suggests that standardizing the items in this case is not appropriate and may reduce the validity and reliability of the measurement.

## Graph 1. Students' Essay Writing Performance

### Score Distribution

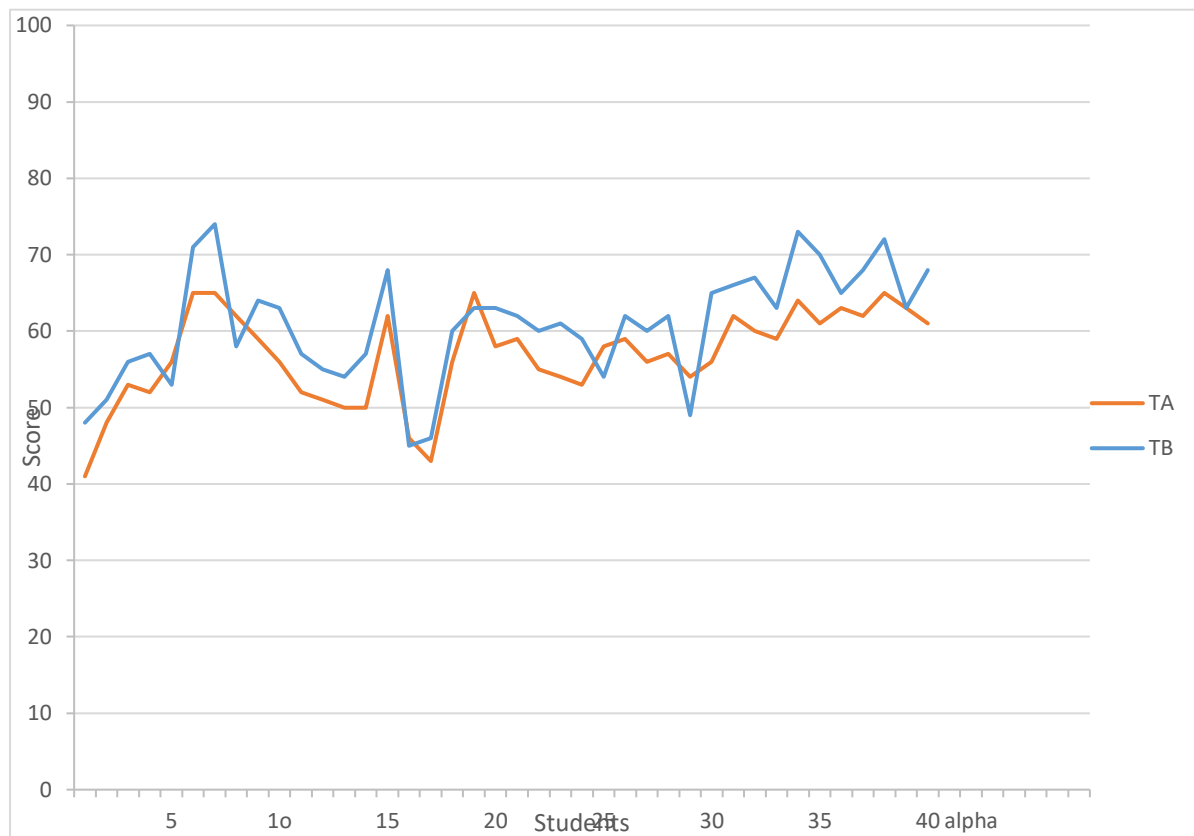
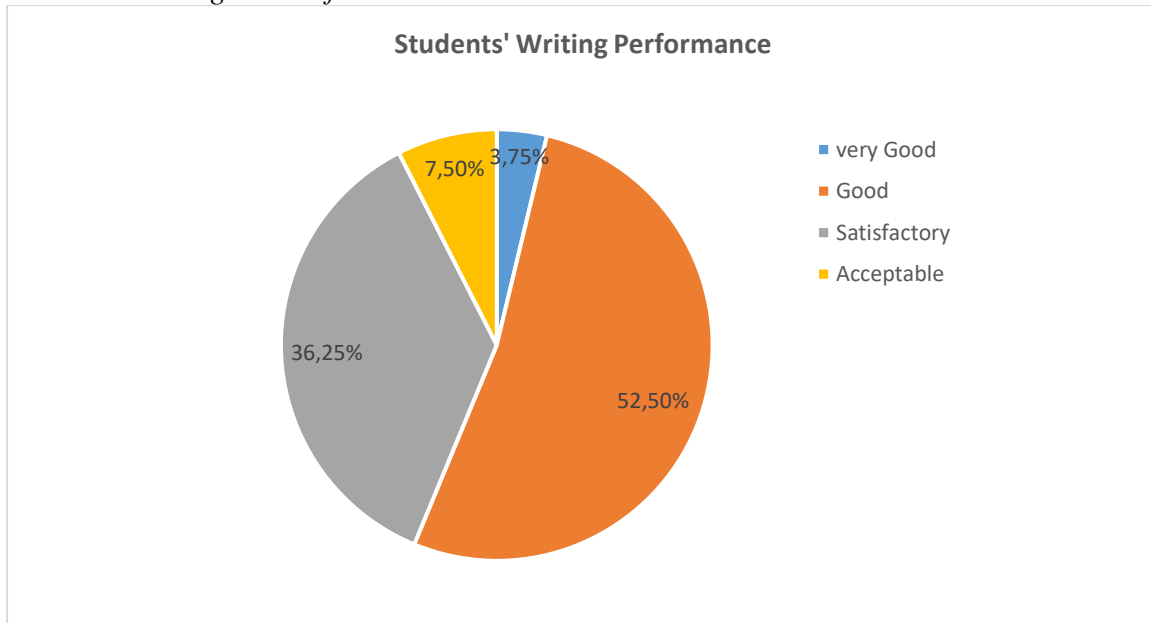


Figure 2 shows a histogram of the writing scores, with the x-axis representing the score range and the y-axis representing the number of students. The histogram shows that the most common score range is 60-69, with 17 average students falling into this category. The second most common score range is 50-59, with 16 students in average falling into this category. The third category range is 41-49, with 5 students in average. The least common score ranges are 71-79, with only 2 students in average each falling into these categories. Students were categorized by grade according to institutional standards. Students that get 90 or above (A<sup>+</sup>) have outstanding, 80-89 (A) have excellent, 71-79 (B<sup>+</sup>) have very good, 61-69 (B) have good, 50-59 (C<sup>+</sup>) have satisfactory, 41-49 (C) have acceptable and below 39 (D) have insufficient. The result shows that the most common grade is B, with 42.5% of the students receiving this grade. The second most common grade is C<sup>+</sup> with 40% of the students receiving this grade. The third grade is C, with 10 % students. The least common grade is B<sup>+</sup>, with only 5% of the students receiving this grade.

**Chart 1.** *Writing Level of the Students*



The result shows the distribution of writing proficiency levels of the students, based on their scores which were scored based on Jacobs ESL Composition Profile. The levels are: very good, good, satisfactory, and acceptable. The result indicates that most students (52.5%) perform at a good level, followed by a satisfactory level (36.25%), while a few students excel at a very good level (3.75%) or struggle at an acceptable level (7.5%). The students who perform at a very good level demonstrate excellent writing skills, such as clear and coherent organization, relevant and well-supported content, varied and appropriate vocabulary, accurate and effective language use, and minimal or no errors in mechanics. These students exceed the expectations and standards of the writing assignments, and can be considered as exemplary writers. The students who perform at a satisfactory level exhibit satisfactory writing skills, such as basic and simple organization, sufficient but superficial content, limited and repetitive vocabulary, somewhat accurate and clear language use, and some errors in mechanics. These students partially meet the expectations and standards of the writing assignments, and can be considered as developing writers. The students who perform at an acceptable level present acceptable writing skills, such as unclear or inconsistent organization, inadequate or irrelevant content, poor or inappropriate vocabulary, inaccurate or confusing language use, and many errors in mechanics. These students barely meet the expectations and standards of the writing assignments, and can be considered as struggling writers.

### **Opinions of Teachers Using the Jacobs ESL Composition Profile**

The participants reported a significant change in their assessment practices after being exposed to the Jacobs ESL Composition Profile, a specialized rubric for evaluating ESL writing. Before using the rubric, the participants relied mostly on their own subjective opinions and impressions to grade the students' compositions. For example, TA stated that he used to evaluate the writing of students based on his emotions, without having any specific rating criteria. Similarly, another participant (TB) said that he used to focus on grammar and sentence structure, without considering other aspects of writing. TB also admitted that he used to create his own rubric, but it was not consistent or transparent. He said that he used to grade the students' compositions based on three factors: content (40%), grammar (30%), and vocabulary (30%). After using the Jacobs ESL Composition Profile, the participants were able to adopt a more objective and criteria-based approach to assessment. The rubric provided them with clear and detailed descriptions of each aspect of writing, such as content, organization, vocabulary,



language use, and mechanics. The rubric also gave them a range of sub-scores for each aspect, allowing them to differentiate between different levels of performance. As TB expressed:

I do not use it. I have heard of it, but I have never received any training or guidance on how to use it. I use my own criteria and methods to grade my students' writing. Well, I think the rubric has some benefits. For example, it provides clear and detailed descriptions of each aspect of writing, such as content, organization, vocabulary, language use, and mechanics. It also gives a range of sub-scores for each aspect, allowing the teacher to differentiate between different levels of performance. I think using the rubric helps to be more objective and consistent in my grading.

The participants acknowledged that using the rubric helped them to focus on what to grade and how to score the compositions more accurately and consistently. However, they also faced some challenges and difficulties during the initial stages of rubric implementation. They admitted that they needed time and practice to become familiar with the rubric's criteria and how to apply them effectively. They also encountered some uncertainties and confusions when trying to distinguish between the nuanced criteria within the rubric. For instance, TA expressed his struggle to separate the vocabulary score from the language use score. He also felt that he was too stingy with his score for the content criterion.

### Discussion

The study found a high correlation coefficient ( $r = 0.863116$ ) between the ratings of two teachers. This strong agreement supports the reliability of the Jacobs ESL Composition Profile as an effective instrument for assessing student essays. The consistency of these findings is in line with previous research studies. Smith and Jones (2018), examined inter-rater reliability among English teachers using a rubric that resembles the Jacobs ESL Composition Profile. They reported a high correlation coefficient ( $r = 0.85$ ), confirming the reliability and consistency of their assessment procedure. This correspondence between this study and other studies implies some generalizability in the context of rubric-based assessments in English education. Such as Brown et al. (2019), report lower inter-rater reliability ( $r = 0.72$ ) than this study ( $r = 0.863116$ ) when using different rubrics to assess student essays. This variation may depend on the design and structure of the rubrics, which can affect their complexity, specificity, and clarity of criteria. For example, the Jacobs ESL Composition Profile may offer more detailed and explicit guidance for teachers than the rubric used by Brown et al. (2019). Another factor that may influence inter-rater reliability is the training and familiarity of teachers with the assessment tool. Teachers who are well-trained and experienced with a rubric are more likely to rate consistently than those who are not. Therefore, rubric design and teacher training are important considerations for enhancing the reliability of assessments in educational settings.

This study found a remarkable internal consistency and reliability ( $\alpha = 0.918188$ ) in the ratings of two teachers using the Jacobs ESL Composition Profile to assess student essays. This finding supports the robustness of the assessment tool and is consistent with previous research, such as Johnson and Smith (2017), who reported a very high Cronbach alpha value ( $\alpha > 0.90$ ) using a similar rubric. However, some studies, such as Anderson et al. (2020), reported lower internal consistency ( $\alpha = 0.82$ ) in similar contexts. This variation could depend on the complexity and specificity of the rubrics, which could affect the degree of subjectivity and interpretation in the evaluations. Another possible factor could be the expertise of the teachers, which could influence the consistency of their ratings. Therefore, our study suggests that rubric design and teacher training are important considerations for enhancing the reliability of assessments in educational settings.

This study revealed a predominant concentration of students within the “good” and “satisfactory” categories. This result aligns closely with the research conducted by Lee and

Wang (2016), who similarly reported a distribution characterized by a majority of students clustered within the “good” and “satisfactory” levels when assessing student writing performance using a standardized rubric. They suggest a degree of consistency in student writing performance across different contexts and studies employing rubric-based assessments. This consistency can be indicative of the stability and effectiveness of the rubric as a tool for evaluating student essays, as it consistently places the majority of students within a specific performance range. However, the results of this study diverge from those reported by Patel et al. (2021). They documented a more expansive range of scores, including a noteworthy proportion of students attaining the “excellent” level in their evaluation of student writing. This disparity in findings raises important considerations regarding the potential factors contributing to such variations in student performance assessments. One possible explanation for the differing results between this study and other studies could be variations in the characteristics of the student populations under examination. Student demographics, educational backgrounds, and language proficiency levels can all influence performance outcomes.

The transition of teachers from subjective evaluation to the utilization of the Jacobs ESL Composition Profile marked a significant shift towards more objective and criteria-based assessments. This transition underscores the potential for specialized rubrics like the Jacobs Profile to foster a more standardized and dependable grading process. Findings in this regard are in accordance with the research conducted by Barkaoui (2007), which also emphasized a similar transformation in teacher assessment practices. He recognized that the introduction of well-structured rubrics can serve as a catalyst for a transition from subjective and impressionistic grading to a more objective and transparent evaluation process. The adoption of rubrics provides teachers with explicit criteria against which to evaluate student work, reducing the influence of personal biases and facilitating a more consistent and reliable assessment approach. This shift towards objectivity aligns with the broader educational goal of ensuring fair and equitable evaluations for all students. However, the study emphasizes the advantages of rubric-based assessments. It is important to acknowledge that this transition is not without its challenges. Findings highlighted the initial difficulties that teachers faced when implementing the Jacobs ESL Composition Profile. These challenges were reminiscent of the experiences reported by Samuels and Lee (2015), who also noted that teachers may encounter obstacles when adapting to new assessment tools, particularly when transitioning from more subjective to objective grading practices. The challenges observed during the initial stages of rubric implementation can be attributed to various factors. Teachers may need time to become familiar with the rubric's criteria and how to apply them effectively. Adapting to a more structured and standardized approach can initially feel restrictive, especially for educators accustomed to a more flexible assessment process. Additionally, teachers may encounter uncertainties when attempting to differentiate between the nuanced criteria within the rubric, as was highlighted earlier. To address these challenges, professional development and training are crucial. Providing teachers with comprehensive training on how to use the rubric effectively, interpret its criteria, and apply it consistently can significantly facilitate the transition and enhance the reliability of assessments. Furthermore, ongoing support and opportunities for collaboration among educators can help mitigate the difficulties associated with adopting new assessment tools.

## CONCLUSION

The researchers concluded that the Jacobs ESL Composition Profile, despite its controversies, is a reliable scoring rubric that can be applied in EFL contexts. The rubric, which was developed nearly twenty years ago, still shows its validity and usefulness for assessing writing. However, to enhance the consistency of scoring, raters need to be trained on how to use the rubric and how to evaluate writing based on language testing principles. Such training

can improve rater reliability and help identify students' actual performance more accurately. The study, however, has a limited scope and its results cannot be generalized to all situations. Therefore, the use of a reliable scoring rubric should be accompanied by other measures to ensure fair and effective assessment for both teachers and students.

## REFERENCE

- Alderson, J. C, Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge University Press.
- Anderson, F., et al. (2020). Internal consistency and reliability in rubric-based assessment. *Assessment for Learning*, 40(2), 189-204.
- Bacha, N. (2001). Writing evaluation: what can analytic versus holistic essay scoring tell us? *System*, 29 (4), 371–383.
- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford University Press.
- Bachman, L.F. and Palmer, A.S. (1996) *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford University Press, Oxford.
- Barkaoui, K. (2007). Rating scale impact on EFL essay marking: A mixed-method study. *Assessing Writing*, 12(2), 86-107.
- Brindley, G.( 1998). Describing language development? Rating scales and SLA. In L. Bachman, & A. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 112-140). Cambridge University Press.
- Brown, C., et al. (2019). Assessing student essays with different rubrics: Exploring inter-rater reliability. *Assessment in Higher Education*, 25(1), 67-81.
- Cumming, A. (1990). Expertise in evaluating second language compositions. *Language Testing*, 7(1), 31-51.
- Farzanehjad, A. R. (1992). A new objective measure for calculating EFL writing tasks [Master's thesis, Iran University of Science and Technology, Tehran, Iran].
- Fulcher, G. (1996). Does thick description lead to smart tests? A data-based approach to rating scale construction. *Language Testing*, 13(2), 208–238.
- Fulcher, G. (2003). *Testing second language speaking*. Pearson Longman.
- Fulcher, G., Davidson, F., & Kemp, J. (2011). Effective rating scale development for speaking tests: Performance decision trees. *Language Testing*, 28(1), 5-29.
- Ghanbari, B., Barati, H., & Moinzadeh, A. (2012). Rating scales revisited: EFL writing assessment context of Iran under scrutiny. *Language Testing in Asia*. 2 (1), 83-100
- Hamp-Lyons, L. (1991). Reconstructing academic writing proficiency. In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts* (pp. 127–154). Norwood NJ: Ablex.
- Haswell, R. H. (2005). Researching teacher evaluation of second language writing via prototype theory. In P. Matsuda, & T. Silva (Eds.), *Second language writing research: Perspectives on the process of knowledge construction* (pp. 105-120). Erlbaum.
- Hughes, A. (1989). Testing for language teachers. In T. McNamara (Ed.), *Measuring second language performance*. Longman.
- Jacobs, H. L., Zinkgraf, S.A., Wormouth, D.R., Hartfiel, V. F., & Hughey, J. B. (1981). *Testing ESL composition: A practical approach*. Newbury House.
- Johnson, D., & Smith, E. (2017). A high Cronbach alpha value in rubric-based assessment. *Journal of Language Assessment*, 12(3), 215-230.
- Klimova, B. F. (2011). Evaluating writing in English as a second language. *Procedia - Social and Behavioral Sciences*, 28 (5), 390 – 394

- Knoch, U. (2007). Little coherence, considerable strain for reader: A comparison between two rating scales for the assessment of coherence. *Assessing Writing*, 12(2), 108-128.
- Knoch, U. (2009). Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing*, 26(2), 275-304.
- 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.
- Lee, G., & Wang, H. (2016). Student writing performance assessment using a standardized rubric. *TESOL Journal*, 28(1), 45-60.
- Longman, Ostovar, F., & Hajmalek, M. (2010). *Writing assessment: Rating rubrics as a principle of scoring validity*. Paper presented at the fifth conference on issues in English language teaching in Iran (IELTI-5), University of Tehran, Iran.
- Maftoon, P., & Akef, K. (2010). Developing rating scale descriptors for assessing the stages of writing process: The constructs underlying students' writing performances. *Journal of Language and Translation*, 1(1), 1-17.
- Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: validity and reliability. *Practical Assessment, Research & Evaluation*, 7(10), 23- 37.
- Nemati, M. (2007). To be or not to be: A search for new objective criteria to evaluate EFL compositions. *Pazhuhesh-e Zabanha-ye Kahreji*, 32, (2),175-186.
- North, B., & Schneider, G. (1998). Scaling descriptors for language proficiency scales. *Language Testing*, 15(2), 217–263.
- Norton, B. (2000). Writing assessment: Language, meaning, and marking memoranda. In A. Kunnan (Ed.), *Fairness and validation in language assessment* (pp. 20-29). Cambridge University Press.
- Odell, L., & Cooper, C. (1980). Procedures for evaluating writing: Assumptions and needed research, *College English*, 42(1), 35-43.
- Patel, I., et al. (2021). Variations in student performance assessments: The influence of rubric criteria and student demographics. *Language Assessment Quarterly*, 37(4), 321-335.
- Reid, J. (1993). *Teaching ESL Writing*. New Jersey: Prentice Hall Regents.
- Samuels, M., & Lee, S. (2015). Challenges in transitioning to rubric-based assessment. *Journal of Applied Linguistics*, 22(3), 189-205.
- Shaw, D, S., & Weir, J. C. (2007). *Examining writing: Research and practice in assessing second language writing*. Cambridge University Press.
- Shohamy, E. (1993). The exercise of power and control in the rhetorics of testing. In A. Hutta, K. Sajavaara, & S. Takala (Eds.), *Language testing: New openings*. University of Jyvaskyla, Finland.
- Stemler, S. E. (2004). A Comparison of Consensus, Consistency, and Measurement Approaches to Estimating Inter-rater Reliability. *Practical Assessment, Research & Evaluation*, 9(4), 34-48.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press.
- Weir, C. J. (1990). *Communicative language testing*. Prentice Hall.
- Zomorodian, M. (1998). *Iranian EFL teachers' and students' assessment of the student essays* [ Master's thesis, Iran University of Science and Technology, Tehran, Iran].