

## Rating scale in writing assessment: Holistic vs. Analytical scales: A review

Alhaji Malah Galti<sup>1</sup>, Sani Saidu<sup>2</sup>, Habu Yusuf<sup>3</sup>, Abba Alkali Goni<sup>4</sup>

<sup>1-3</sup>Department of English Language- Yobe State University, Damaturu, Nigeria

<sup>4</sup>Federal College of Freshwater Fisheries Technology Baga, Nigeria

### Abstract

Language learning could be best measured through testing and assessment. Teachers appreciate learners' learning strengths and weaknesses through testing. This may assist teachers in making judgement about the learning process in general and learning performance in particular. Therefore, how to conduct the testing is of paramount to experts and practicing teachers. The processes involved in testing, especially the choice of scoring scale in the testing process depends on what needed to be tested. Which scoring scale is appropriate to be employed to test writing task is a great concern to experts. Basically, language writing raters used any of the following rating scales in assessment- Holistic or analytical scales. This paper reviewed these scoring scales and identified why and when to employ these scales. In addition, the pros and cons of each of the scale were presented. Finally, this paper suggested that holistic scale should be employed for beginners when focusing on the learners' ability to develop content; while the analytic scale should be considered for learners with higher or intermediate level and when the focus is on mechanical accuracy.

**Keywords:** analytic, holistic, rater, scoring scale, test reliability, test validity, testing

### 1. Introduction

Test, generally involves the measuring of learners' behaviour after learning session. Language test could be said to be making judgement on learners' knowledge about a language and their ability to use the language. The main purpose of a test is to see what learners have learnt. This will not only benefit the learners in understanding their strength and weakness but also assist the teachers in making judgement and decision about their methods, students and materials used. Language test is needed to demonstrate one's competence in the use of language. Because assessment generally is considered as one of the programs of teaching and learning (Pandiya, 2013) <sup>[12]</sup>.

How to assess and measure writing competence is of interest to experts that required to be explored. It was reported that one important aspect of measuring writing ability is the use of a valid and a reliable rating or grading scale (Charney, 1984a) <sup>[2]</sup>. However, the choice of appropriate rating scale is a challenging task. In the recent years, pronouncing judgement on a writing text has received a significant attention (Saaty, 2008) <sup>[14]</sup>. Therefore, the question of what type of scoring scale should be employed is another issue raised. Do the scoring scale applied in the assessment of writing really satisfied the expected purpose of writing assessment (measuring the writers' actual ability)? In writing assessment, the appropriate rating scale used is important for making judgement about the writers' ability. McNamara (1996) <sup>[11]</sup> illustrated in diagram the main purpose of assessment as the central point. According to him, a valid assessment should target the central point (performance).



**Fig 1:** Purpose of assessment

The rationale for using particular rating scale is related to the output, writers' level, writing purpose, and for making judgement. And this will satisfied construct validity. Similarly, it is pertinent to note that a valid writing assessment should consider the true ability of the learner, the purpose of the writing as well (Charney, 1984a) <sup>[2]</sup>. Because validity and reliability in language test are ominous. Validity refers to what extent the test really measures what it claims to measure (Iwashita & Grove, 2003) <sup>[9]</sup> while reliability is the consistency of a measure (Charney, 1984a; C. R. Cooper & Odell, 1977) <sup>[2, 5]</sup>. Ghanbari, Barati, and Moinzadeh (2012) citing the work of Weigle, (2002) who introduced five factors to be considered

when developing rating scale in writing assessment. These include.

- What type of rating scale should be used?
- Who is going to use the rating scale and what is the purpose of the scale?
- What aspect of writing are most important and how would they be divided up?
- What will the descriptors look like and how many scoring levels will be used?
- How will scores be reported?

In this paper, therefore, we will discuss the two types of rating scale in writing assessment, which is one of the conditions to be considered when developing rating scale. Moreover, when and why to employ a particular scale will also be discussed.

## 2. The holistic scoring

Holistic Rating Scales (also “impressionistic”) are a type of scoring that involves assigning of the score to the whole text (Knoch, 2011) <sup>[10]</sup>. In other word, assigning of a single score to a whole piece of writing on the basis of an overall or general impression of the text. The main focus of such type of scoring (grading) in writing is on how the message or the idea being conveyed in the work of writing (Terry, 1989) <sup>[15]</sup>. Much attention is not given to the individual parts such like grammatical accuracy, spellings and the use of tenses but the message.

Holistic scoring encourages the development of ideas in writing (especially, in CLT approach class). Learners’ creativity and development of writing idea is what is considered when assessing writing (Terry, 1989) <sup>[15]</sup>. Because the assessing writing skill should go beyond mastery of convention (grammar rules) (Charney, 1984a) <sup>[2]</sup>. Here, the raters taking into consideration the impact the errors in the text have on the overall comprehensibility of the piece (text, produced) rather than focusing mainly about the grammar, devices used, or even length of paragraph etc. C. R. Cooper and Odell (1977) <sup>[5]</sup> in the work of (Barkaoui, 2007) <sup>[1]</sup> described holistic scale to includes “any procedure which stop short of enumerating linguistic, rhetoric, or information features of a piece of writing” (p. 4).

One of the best advantages of this scoring is that it rates the piece based on the writer’s level and the purpose of instruction (C. R. Cooper, 1977) <sup>[5]</sup>. Similarly, Barkaoui (2007) <sup>[1]</sup> maintains that, though, holistic scale is weak in reliability, but has high validity because it assess the overall writing proficiency of the writer. This means that construct validity is meet since holistic is concerned with what skill or ability is to be measured. The raters rate the quality of the writing. The grading style in holistic scoring, for example, rate writing using the following scale.

1. Excellent (writing) should demonstrate strong control of the language; proficiency and a variety of grammatical usage with few significant errors; broad command of vocabulary and of idiomatic language.
2. Very good (writing) should demonstrate a good general control of grammatical structures despites any errors and/or some awkwardness of style. Good use of idioms and vocabulary. Read smoothly overall.
3. Good (writing) should demonstrate a fair ability to express ideas in the target language; correct use of simple

grammatical structures or use of more complex structures without numerous serious errors. Some apt vocabulary and idioms. Occasional signs of fluency and sense of style.

4. Fairly good (writing) should demonstrate a weak use of language with little control of grammatical structures. Limited vocabulary. Frequent use of L1. Occasional redeeming features.
5. Fair (writing) should demonstrate clearly unacceptable from most points of view. Almost total lack of vocabulary resources, little or no sense of idiom and/or style. Essentially translated from L1.
6. Poor (writing) is full of floating points. A one point bonus should be awarded for a coherent and well organized essay or for a particularly inventive one

(Hughes, 1989:86) <sup>[8]</sup>

This scoring scale can be a good rating scale for beginners of ESL writers. Another rationale of this rating scale is for language graders to focus on the development of the thinking skills in beginning writers and build quest for writing, especially when they were not heated with too much superficial features of their writing while grading (C. R. Cooper & Odell, 1977) <sup>[5]</sup>. Cooper & Odell (1997) <sup>[5]</sup> gave the definition of competence in writing, which need to be developed first in beginners, as “ the ability to discover what one wishes to say and convey one’s message through language (in writing), syntax and content that are appropriate for one’s audience and purpose (p 103). Therefore, when assessing writing, the presentation of fact, ideas and message should be the focus while paying attention to certain errors and their impact on the comprehensibility of the piece (text).

Raters using this scoring type are not editing or diagnosing the weakness of a text in terms of it grammar and use of convention in writing but rather it is a set of procedure for assigning value to a piece of writing according to an established criteria (Charney, 1984) <sup>[2]</sup>. Agreeing with this, Barkaoui (2007) <sup>[1]</sup> states that rating scale is generally developed to target test-takers with particular features who aim to accomplish specific goal in specific contexts. Therefore, assessing writing should go beyond the mastery of the superficial features of the writing but as Hirsch, (1977) cited by (Ruegg & Sugiyama, 2013) <sup>[13]</sup> assert that assessment should include the relative readability and comprehensibility of the message. By relative readability of a text, Hirsch explained it as a standard, which measures how well ideas are presented. Affirming, Hughes (1989) <sup>[8]</sup> confirmed that the whole (of a text) is more often greater than the sum of its parts in a writing.

## 3. Reliability of holistic scale

- Raters should be trained to conform to agree upon criteria of judgment and be consistent on their judgement.
- They must be language experts and come from similar background with the test takers
- And the rating guide must be appropriate

## 4. Validity

For holistic measurement scale to be valid, the purpose of the task should be spelt out clearly and focused mainly on it. Secondly, proficiency level should be considered and measure

their ability only. For the holistic to be valid, it must only measure the salient skill of conveying and how errors in the writing affect the whole text rather mastery of writing convention. However, there is strong debate on the acceptable and valid criteria for measuring writing task except for an attempt by Hirsch's (1977)<sup>[5]</sup> "relative readability" as reported by (Ruegg & Sugiyama, 2013)<sup>[13]</sup>. It a standard which measures how well ideas are presented. He believes that writer's ability to present ideas remains constant across writing task.

### 5. The analytical scoring

Analytic Rating Scales are rubrics that include explicit performance expectations for each possible rating, for each criterion. Analytic rating scales are especially appropriate for complex learning tasks with multiple criteria. On the other hand, it involves the scoring of the individual parts that constitute the text. The attention is about the features and elements that form the text such like grammatical accuracy, use of tense, spellings, text organization- coherence and cohesion, paragraphing, the use of idioms etc. Here, the major concern of the assessment is on the individual element and salient features of the text. As Charney (1984b)<sup>[3]</sup> in his study stated what need to be tested to include the spelling, the use of punctuation, sentence structure and order and some aspect of style. Equally observed by Hughes (1989)<sup>[8]</sup> in scoring text using analytical scale, each element or components is given equal weight. But, too much concentration on the elements and features of the piece to ensure accuracy may divert attention from the overall effect of the piece.

Organization of the text is highly valued by analytical raters. Ruegg and Sugiyama (2013)<sup>[13]</sup> citing (Harris, 1977; Freedman, 1979) reported that analytical raters consider organization to be of great importance in academic writings. Therefore, Charney (1984a)<sup>[3]</sup> explained what is organization in writing. It include essay structure, paragraph structure, use of cohesive devices and coherence. Similarly, (Ruegg & Sugiyama, 2015) in their study "Organization of ideas in writing: what are raters sensitive to?" gave paragraphing, the use of cohesive devices and coherence as writing organization. In this study, they evaluated in each text (essay), the number of paragraphs, the coherence of information within paragraphs and the number of the cohesive devices used. They classified the number of paragraphs and the number of the cohesive devices as superficial physical features and coherence as deeper textual characteristics (see appendix i. for the example of analytical scoring guide by raters while rating writing text and classification of the individual components as Culled from KEPT Essay Rating Scales).

### 6. Conclusion and Suggestions

In conclusion, teachers need to understand that students writing can be evaluated on five product factors which includes fluency, content, convention, syntax and vocabulary (Pandiya, 2013)<sup>[12]</sup>. It is also important to bear in mind that writing can be assessed for different purposes to measure the students writing performances at different level of writing skill development. Therefore, teachers' role is to create opportunity for students to write and encourage them to do so. And this can be achieved with the appropriate feedback to the

writers' work. The teacher can facilitate the development of more skilful writing techniques at a later time. This can be achieved with careful selection of an appropriate rating scale at different level for different purposes. As Cooper, (1984)<sup>[6]</sup> proposes the assessment of the beginner writers should be by holistic scoring where the focus is on the content development. Analytic scoring should be applied to higher level of writer and as well when concentrating on the superficial features but definitely not to assess the writer writing ability at their primary writing level. With these conclusion, this paper encourage teachers and scholars in Language testing and evaluation to explore the type of language rating employed by the teachers of English as a second language.

### 7. References

1. Barkaoui K. Rating scale impact on EFL essay marking: A mixed-method study. *Assessing writing*. 2007; 12(2):86-107.
2. Charney D. The validity of using holistic scoring to evaluate writing: A critical overview. *Research in the Teaching of English*, 1984a, 65-81.
3. Charney D. The Validity of Using Holistic Scoring to Evaluate Writing: A Critical Overview. *Research in the Teaching of English*. 1984b; 18(1):65-81.
4. Cooper CR. Holistic evaluation of writing. *Evaluating writing: Describing, measuring, judging*, 1977, 3-31.
5. Cooper CR, Odell L. *Evaluating writing: Describing, measuring, judging*, 1977.
6. Cooper PL. The assessment of writing ability: A review of research. *ETS Research Report Series*. 1984; 1984(1):1-46.
7. Ghanbari B, Barati H, Moinzadeh A. Rating scales revisited: EFL writing assessment context of Iran under scrutiny. *Language Testing in Asia*. 2012; 2(1):83.
8. Hughes A. *Testing for language teachers*. Cambridge: Cambridge University Press, 1989.
9. Iwashita N, Grove E. A comparison of analytic and holistic scales in the context of a specific-purpose speaking test, 2003.
10. Knoch U. Rating scales for diagnostic assessment of writing: What should they look like and where should the criteria come from? *Assessing writing*. 2011; 16(2):81-96.
11. McNamara TF. *Measuring second language performance*: Addison Wesley Longman, 1996.
12. Pandiya A. Rubrics on Scoring English Tests for Four Language Skills. *Ragam*. 2013; 13(1).
13. Ruegg R, Sugiyama Y. Organization of ideas in writing: what are raters sensitive to? *Language Testing in Asia*. 2013; 3(1):8.
14. Saaty TL. Decision making with the analytic hierarchy process. *International journal of services sciences*. 2008; 1(1):83-98.
15. Terry RM. Teaching and evaluating writing as a communicative skill. *Foreign Language Annals*. 1989; 22(1):43-52.