Assessing Writing: A Review of the Main Trends

Mohammad Ali Salmani Nodoushan
(dr.nodoushan@gmail.com)

Iranian Institute for Encyclopaedia Research
Ministry of Science, Research and Technology, Tehran, IRAN

Abstract
As a language skill, writing has had, still has and will continue to have an important role in shaping the scientific structure of human life in that it is the medium through which scientific content is stored, retained, and transmitted. It has therefore been a major concern for writing teachers and researchers to find a reliable method for evaluating and ensuring quality writing. This paper addresses the different approaches to scoring writing and classifies them into a priori *scoring systems (including holistic and analytic scoring), and* a posteriori *trait-based scoring systems (including primary-trait and multiple-trait scoring).*

Keywords: Writing, Assessment, Evaluation, Holistic scoring, Trait-Based scoring, Primary-Trait scoring, Multiple-Trait scoring.

1. INTRODUCTION

This paper is an attempt at summarizing the literature on the assessment of writing. First, the term assessment is defined. Then, it is related to evaluation. Next, the paper adopts a narrow perspective and focuses on the skills of writing. Different approaches to the assessment of (student) writing will then be described.

2. WHAT IS ASSESSMENT?

In its most fundamental sense, assessment aims at supporting and improving student learning. Assessment, as a term in the academic community, stems from a movement towards "accountability". It originates from the conflict between a "traditional view" of what teachers need to do and a "concern" for what learners can and do actually learn. The traditional view is referred to as the "inputs" view and the latter—the concern—is called the "outputs" view.

Whenever information is collected with the purpose of guiding future instruction, it can be called assessment (Peha, 2011). An example could be a statement like this:

"When I looked at their last published pieces, I noticed that many kids were having trouble with run-on sentences" (Peha, 2011, p. 29).

Good assessment requires at least two main considerations:

1) It uses specific and appropriate language to describe the data gathered and the patterns that are observed.

2) It is based on authentic data gathered in an authentic manner from within an authentic context.

In doing any assessment, the teacher should ask himself the following question:

• How do I plan to use the data I am gathering to guide my instructions?

Therefore, assessment has to do with what students know, what they are able to do, and what values they have when they leave school. It is concerned with the overall and collective impact and influence of a program on student learning (Peha, 2011).

3. THE LINK BETWEEN ASSESSMENT AND EVALUATION

Assessment is closely related to evaluation. Peha (2011) defines evaluation as any *decision* that is made based upon the information which has already been gathered through assessment. An example of evaluation could be the following:

Because I noticed that many kids were not using periods and capitals correctly, I'll teach some sentence punctuation mini-lessons in writing and support that with simple inquiry activities during reading time where I'll have the kids identify sentence boundaries by ear using expressive reading techniques (Peha, 2011, p. 29).

Evaluation is, therefore, a tool that can be used to help teachers judge whether a teaching program or a classroom approach is being used as it was planned to be; it is also a means to assess the extent to which stated goals and objectives are being achieved (McLaughlin, 1975). Evaluation allows teachers to answer the following questions:

• Are we doing for our students what we said we would?

• Are students learning what we set out to teach?

• How can we make improvements to the curriculum and/or teaching methods?

A good evaluation has the following characteristics:

1. It includes a specific plan of action.

2. It uses the assessment data as its rationale.

With this short description of assessment and evaluation, this paper will now move on to a description of writing followed by a consideration of the main trends in the assessment of writing.

4. WRITING

Writing is a language skill that has attracted the attention of many language schools and institutes. Language skills have traditionally been classifies as receptive and

productive. Receptive skills are those in which the individual receives language produced by others. They include reading and listening. On the contrary, productive skills include speaking and writing. They are two critical skills which form a main component of the complex process of communication (Hyland, 2003). There are many different reasons for communication between individuals. For example, individuals may have something they wish to express either verbally or in writing. There may even be something (either verbally or in writing) that individuals wish to receive or learn.

Writing, as a productive skill, requires a great degree of accuracy. Many language teachers agree that writing is in many ways the most difficult language skill to learn in comparison to other language skills (Hyland, 2003). It is therefore the most difficult language skill to teach, and even to assess. Needless to say, fostering useful and effective language skills in students is a painstaking task if the language teacher lacks enough experience and fails to provide appropriate practice (Kroll, 1990). When it comes to writing, the language teacher's job is even more difficult. Developing writing requires the teachers' use of controlled lessons, authentic tasks, and real-life experiences (Swales & Feak, 1994).

Since any teaching activity, especially in a formal setting, is followed by an assessment activity, the teaching of writing, too, requires an assessment phase. The aim of the assessment phase is to provide information on both the degree to which students have achieved and the extent to which the teaching program has been useful (Hyland, 2003). In the past few decades, several approaches to the assessment of writing have emerged. In the following sections, these approaches are described.

5.   ASSESSMENT OF WRITING[1]

Over the past few years language testing specialists have called for performance assessment in EFL contexts. Advocates of performance assessments maintain that every task must have performance criteria for at least two reasons. On the one hand, the criteria define for students and others the type of behavior or attributes of a product which are expected. On the other hand, a well-defined scoring system allows the teacher, the students, and others to evaluate a performance or product as objectively as possible. If performance criteria are well defined, another person acting independently will award a student essentially the same score. Furthermore, well-written performance criteria will allow the teacher to be consistent in scoring over time. If a teacher fails to have a clear sense of the full dimensions of performance, ranging from poor or unacceptable to exemplary, he or she will not be able to teach students to perform at the highest levels or help students to evaluate their own performance (Hyland, 2003).

In developing performance criteria, one must both define the attribute(s) being evaluated and also develop a performance continuum. For example, one attribute in the evaluation of writing might be writing mechanics, defined as the extent to which the student correctly uses proper grammar, punctuation, and spelling (Birjandi, Alavi & Salmani Nodoushan, 2004). As for the performance dimension, it can range from high

---

[1]Much of the information presented in this and subsequent sections is based on Hyland (2003).

quality (well-organized, good transitions with few errors) to low quality (so many errors that the paper is difficult to read and understand).

Testers and teachers should keep in mind that the key to developing performance criteria is to place oneself in the hypothetical situation of having to give feedback to a student who has performed poorly on a task. Advocates of performance assessment suggest that a teacher should be able to tell the student exactly what must be done to receive a higher score. If performance criteria are well defined, the student then will understand what he or she must do to improve. It is possible, of course, to develop performance criteria for almost any of the characteristics or attributes of a performance or product. However, experts in developing performance criteria warn against evaluating those aspects of a performance or product which are easily measured. Ultimately, performances and products must be judged on those attributes which are most crucial (Hyland, 2003).

Developing performance tasks or performance assessments seems reasonably straightforward, for the process consists of only three steps. According to Hyland (2003), the reality, however, is that quality performance tasks are difficult to develop. With this caveat in mind, the three steps include:

1. Listing the skills and knowledge the teacher wishes to have students learn as a result of completing a task. As tasks are designed, one should begin by identifying the types of knowledge and skills students are expected to learn and practice. These should be of high value, worth teaching to students and worth learning. In order to be authentic, they should be similar to those which are faced by adults in their daily lives and work;

2. Designing a performance task which requires the students to demonstrate these skills and knowledge. The performance tasks should motivate students. They also should be challenging, yet achievable. That is, they must be designed so that students are able to complete them successfully. In addition, one should seek to design tasks with sufficient depth and breadth so that valid generalizations about overall student competence can be made;

3. Developing explicit performance criteria which measure the extent to which students have mastered the skills and knowledge. It is recommended that there be a scoring system for each performance task. The performance criteria consist of a set of score points which define in explicit terms the range of student performance. Well-defined performance criteria will indicate to students what sorts of processes and products are required to show mastery and also will provide the teacher with an "objective" scoring guide for evaluating student work. The performance criteria should be based on those attributes of a product or performance which are most critical to attaining mastery. It also is recommended that students be provided with examples of high quality work, so they can see what is expected of them.

6. APPROACHES TO SCORING WRITING

Scoring writing is a very delicate task. There is still a lot of controversy among teachers as to how students' writing assignments should be scored. Traditionally a

student's writing performance was judged, in a norm-referenced approach, in comparison with the performance of others. Over the past few decades, however, this norm-referenced method has largely given way to criterion-referenced procedures. In a criterion-referenced approach to scoring writing, the quality of each essay is judged in its own right against such external criteria as coherence, grammatical accuracy, contextual appropriateness, and so on. According to Hyland (2003), such an approach takes a variety of forms and falls into three main categories: (a) holistic, (b) analytic, and (c) trait-based. As Weigle (2002) claims, the holistic approach offers a general impression of a piece of writing; the analytic approach is based on separate scales of overall writing features; and the trait-based approach takes a particular task into consideration and judges performance traits relative to its 'trait' requirements (Hyland, 2003).

6.1 Holistic scoring

A holistic scale is based on a single, integrated score of writing behavior. The aim of this method is to rate a writer's overall proficiency. To this end, a general and often individual impression of the quality of a writing sample is made. This approach to scoring students' written performances is global and tacitly reflects the idea that "writing is a single entity which is best captured by a single scale that integrates the inherent qualities of the writing" (Hyland, 2003, p. 227). The holistic approach stands in sharp contrast to earlier methods of writing assessment where the rater/teacher tried to find errors in students' writing—usually through the 'red-pen' method (Salmani Nodoushan, 2007a). As White (1994, cited in Hyland, 2003) suggests, the holistic approach pinpoints and emphasizes what writers 'can do well' rather than identifying writers' incompetence in writing and their deficiencies. Holistic scoring is relatively easy to use, but this approach to scoring writing is quite short-sighted in that it reduces writing to a single score. It is rather impressionistic and fails to pay attention to details by providing a score for each of them. As such, holistic scoring prevents teachers from gaining any diagnostic information which is crucial for subsequent remedial teaching. The holistic scoring approach also has certain connotations for training raters; raters must be carefully trained to respond in the same way to the same features in different students' writings because the holistic approach requires a response to the text as a whole. Cohen (1994, p. 317, cited in Hyland, 2003) summarizes the advantages and disadvantages of the holistic method as follows:

Table 1. Advantages and disadvantages of the Holistic Method
(based on Cohen, 1994, p. 317)

Advantages	*Disadvantages*
Global impression not a single ability	Provides no diagnostic information
Emphasis on achievement not deficiencies	Difficult to interpret composite score
Weight can be assigned to certain criteria	Smoothes out different abilities in sub-skills
Encourages rater discussion and agreement	Raters may overlook sub-skills
	Penalizes attempts to use challenging forms
	Longer essays may get higher scores

	One score reduces reliability
	May confuse writing ability with language proficiency

Hyland (2003) further notices that the reliability of scores gained through the holistic approach improves when two or more trained raters score each paper. Without guidance, however, raters are prone to trouble and error in that they will find it difficult to agree not only on the specific features of good writing but also on the relative quality of the papers they are asked to rate. Nevertheless,young teachers gradually gain the experience that will lead them to develop the confidence and skill which will enable them to score students' writing consistently.

According to Hyland (2003), scoring rubrics or guides can be used which will help teachers/raters.Such scoring rubrics or guides are quite often bands of descriptions which correspond to particular proficiency or rhetorical criteria. Hyland also notices that scoring rubrics are commonly designed to suit different contexts; rubrics seek to reflect the goals of the course and describe what writing teachers consider as good writing. This, of course, requires that scoring rubrics be written in such a careful and precise way as to avoid ambiguity.

One possibility in writing precise scoring rubrics for writing is to make sure that the rubrics will have multiple-step (e.g., nine- or ten-step) scales. This should not misguide the writing teacher to think that a greater number of steps will correspond to a more precise scoring rubric;on the contrary, it is unlikely that scorers can reliably distinguish more than about nine bands (Hyland, 2003). It is on this ground that most holistic rubrics found in the literature on writing assessment have between four to six bands. Examples of holistic rubrics can be found in Cohen (1994), Hamp-Lyons (1991), and White (1994). The following sample rubric for a holistically-scored essay can be found in Hyland (2003, p. 228).

Table 2. Sample rubric for a Holistically-Scored essay
(adopted from Hyland, 2003, p. 228)

Grade	Characteristics
A	The main idea is stated clearly and the essay is well organized and coherent. Excellent choice of vocabulary and very few grammatical errors. Good spelling and punctuation.
B	The main idea is fairly clear and the essay is moderately well organized and relatively coherent. The vocabulary is good and only minor grammar errors. A few spelling and punctuation errors.
C	The main idea is indicated but not clearly. The essay is not very well organized and is somewhat lacking in coherence. Vocabulary is average. There are some major and minor grammatical errors together with a number of spelling and punctuation mistakes.
D	The main idea is hard to identify or unrelated to the development. The essay is poorly organized and relatively incoherent. The use of vocabulary is weak and grammatical errors appear frequently. There are also frequent spelling and punctuation errors.
E	The main idea is missing and the essay is poorly organized and generally incoherent. The use of vocabulary is very weak and grammatical errors appear very frequently. There are many spelling and punctuation errors.

It should be noted that a single rubric cannot and should not be used for scoring all forms of writing regardless of their degree of complexity; rather,it is both possible and desirable to devise more complex rubrics for complicated forms of writing. Devising complex rubrics will of course require attention to the complexity of the writing task, its genre, and its topic (Hyland, 2003). Other considerations that can be taken into account in devising holistic-scoring rubrics include the fact that students may have to express and counter different viewpoints, and that they may have to draw on suitable interpersonal strategies. In discussing this point, Hyland notices the existence of a dilemma: while more delicate holistic rubrics are feasible, they are also more difficult to apply since the rater may encounter texts which simultaneously display characteristics from more than one category (Hyland, 2003). As such, rubrics have to be devised on the basis of the criterion of 'optimality' which will result in the development of an optimal set of rubrics clearly defining separate sets of features for which each piece of writing is to be scored.

It should be noted that, as Hyland (2003, p. 228) puts it, even the above simple rubric may fail to provide an obvious basis for scoring "where, for instance, a text has a clear thesis statement and displays appropriate staging for the genre but contains numerous significant grammatical errors, so that features from B and C grades overlap". In such a situation raters may choose to make finer distinctions with + and − subdivisions (i.e., grading the problematic writing as a B − or a C+).

6.2   Analytic scoring

Analytic scoring was suggesting in response to the inherent flaw in holistic scoring: that features of good writing should not be collapsed into one single score. Raters who employ analytic scoring procedures often judge a written text against a carefully-devised set of criteria important to good writing. Features of good writing are classified into certain separate categories, and raters must give a score for each category. This helps ensure that features of good writing are not collapsed into one single overall score, and, as such, provides more information than a single holistic score could ever do. In other words, analytic scoring procedures more clearly define the features to be assessed by separating, and sometimes weighting, individual components. This scoring procedure is, therefore, more effective in discriminating between weaker texts. Analytic scoring rubrics are in wide use today, and have separate scales for content, organization, and grammar; scales for vocabulary and mechanics are sometimes added separately. Each of these parts is assigned a numerical value (Hyland, 2003).

The idea behind analytic scoring is that writing quality is not a holistic unified scale; rather, it is composed of certain separate features; as such, the tacit assumption which underlies analytic methods of scoring is to encourage teachers to pay close attention to the specific features of writing quality captured in the rubrics for analytic scoring. Analytic scoring rubrics assist rater training, and give more detailed information; they are also useful as diagnostic and teaching tools. Through the implementation of analytic scoring rubrics, writing teachers will be able to pinpoint weaknesses in students' writings which can then be followed up by remedial instructions (Salmani Nodoushan, 2007b). Hyland (2003) suggests that it is recommended that raters,

when devising an analytic rubric, use explicit and comprehensible descriptors that relate directly to what is taught. This allows teachers to target writing weaknesses precisely. It also provides a clear framework for feedback, recast, and revision. The criteria delineated in an analytic rubric can be introduced early in the writing course to show students how their writing will be assessed. They can also give the students an understanding of writing properties and features which their teachers will value in their writings.

Like holistic scoring, analytic scoring, too, is not without its shortcomings. As Hyland (2003, p. 229) noticed, some critics of analytic scoring procedures "point to the dangers of the halo effect; results in rating one scale may influence the rating of others, while the extent to which writing can be seen as a sum of different parts is controversial". Cohen (1994) and McNamara (1996) have identified the advantages and disadvantages of analytic rubrics as follows:

Table 3. Advantages and disadvantages of Analytic Rubrics
(based on Cohen (1994) and McNamara (1996))

Advantages	*Disadvantages*
Encourages raters to address the same features	May divert attention from overall essay effect
Allows more diagnostic reporting	Rating one scale may influence others
Assists reliability as candidate gets several scores	Very time consuming compared with holistic method
Detailed criteria allow easier rater training	Writing is more than simply the sum of its parts
Prevents conflation of categories into one	Favors essays where scalable info is easily extracted
Allows teachers to prioritize specific aspects	Descriptors may overlap or be ambiguous

6.3   Trait-Based scoring

Both analytic and holistic scoring were a priori in that they assumed a pre-determined set of criteria which could distinguish good writing from poor writing, and according to which each piece of writing could be evaluated. A tacit assumption behind both analytic and holistic scoring is that writing is not context-sensitive; however, trait-based approaches to scoring writing are context-sensitive and, as such, differ from both holistic and analytic scoring methods. They do not presuppose that the quality of a text can be based on a priori views of good writing (Hyland, 2003). Rather, as Hamp-Lyons (1991) claims, trait-based instruments are designed to clearly define the specific topic and genre features of the task being judged. The goal that trait-based scoring approaches is to create criteria for writing unique to each prompt and the writing produced in response to it. Trait-based approaches are therefore task-specific.

As Hyland (2003) suggests, trait-based approaches fall into two main categories: (a) primary-trait scoring, and (b) multiple-trait scoring. The following sections provide a separate description of each scoring system:

6.3.1 Primary-Trait scoring

Primary-trait scoring is in some way similar to holistic scoring in that in primary-trait scoring, too, one score is assigned to the criteria intended for scoring; however, it differs from holistic scoring in that the criteria intended for scoring a piece of writing are sharpened and narrowed to just one feature relevant to the writing task in question (Hyland, 2003). This scoring system defines a primary trait in the writing task which will then be scored. Very often a critical feature of the writing task is considered to be the primary trait, and that feature is what will be scored. Examples of primary traits to be scored include appropriate text staging, creative response, and effective argument, reference to sources, audience design, and so forth. Genre-based approaches to scoring writing, for instance, may address the correct sequencing of rhetorical moves in a piece of writing as the primary trait for which scores will be assigned. The rater will then evaluate the written text to see if the rhetorical moves in the text have been sequenced correctly or not, and the text will be scored accordingly.

One shortcoming of primary-trait approaches is that it is not possible to respond to everything at once. In practice, primary-trait raters quite often find it hard to focus exclusively on the specified trait in focus; they may unknowingly include other traits in their scoring. Another shortcoming of this scoring system is its lack of generalization. A necessary consideration for primary-trait scoring is that a very detailed scoring guide needs to be devised for each specific writing task. This limits the scoring system in that it can only be practically used in courses where teachers need to judge learners' command of specific writing skills rather than more general improvement (Hyland, 2003).

6.3.2 Multiple-Trait scoring

Multiple-trait scoring is very much similar to analytic scoring. Here, too, several features in the writing task will be scored. While analytic scoring employs a pre-defined set of features to be scored (i.e., it is a priori), multiple-trait scoring is task-specific, and the features to be scored vary from task to task. This requires that raters provide separate scores for different writing features. Since each writing task has a specific set of writing features that are relevant to it, multiple-trait raters are expected to ensure that the features being scored are the features relevant to the writing assessment task at hand. It is not surprising, therefore, that many raters find multiple-trait scoring as the ideal scoring procedure for writing tasks.

Multiple-trait scoring, as Hyland (2003, p. 230) puts it, *"treats writing as a multifaceted construct which is situated in particular contexts and purposes, so scoring rubrics can address traits that do not occur in more general analytic scales"*. The examples Hyland (ibid) provides include the ability to *"summarize a course text"*, *"consider both sides of an argument"*, or *"develop the move structure of an abstract."*

Multiple-trait scoring is very flexible because each task can be related to its own scale; the scoring system can then be adapted to the context, purpose, and genre of the elicited writing. Due to its task-specific nature, multiple-trait scoring has clear benefits for raters, students, and course designers. It encourages raters to attend to 'relative'

strengths and weaknesses in an essay. As for the students, it provides opportunities for them to have access to detailed feedback in relation to their writing performance; in other words, teachers can use multiple-trait scoring to identify students' weaknesses and to provide them with appropriate feedback and remedial instructions. Multiple-trait scoring also assists wash-back into instruction directly—what is commonly known as remedial instruction (Salmani Nodoushan, 2007b). Multiple-trait scoring, therefore, provides rich data which will inform decisions about remedial instruction and course content. One major disadvantage of multiple-trait scoring is that it requires enormous amounts of time to devise and administer. Another major disadvantage is that teachers may still fall back on traditional general categories in their scoring although traits are specific to the task (see Cohen, 1994, p. 323).

7. CONCLUSIONS

Writing, as a productive skill, is perhaps the most difficult language skill to teach, and the most delicate to assess. Based on the discussion presented above, it can be concluded that the move towards a reliable scoring system for students' writing performance has resulted in the emergence of task-specific scoring system that address writing features specific to each writing task. The move has been from *a priori* scoring systems (i.e., analytic and holistic) to *a posteriori* ones (primary-trait and multiple-trait). It was also noticed in the paper that, when faced with the mental requirements of *a posteriori* scoring systems, teachers may fall back on the traditional *a priori* scoring systems. It must be noted that, while the multiple-trait scoring approach is perhaps the most popular one today, research on writing will definitely open new avenues in the future.

REFERENCES

Birjandi, P., Alavi, S. M., & Salmani Nodoushan, M. A. (2004). *Advanced writing*. Tehran: Zabankadeh Publications.

Cohen, A. D. (1994). *Assessing language ability in the classroom*, 2nd ed. Boston: Newbury House / Heinle & Heinle.

Hamp-Lyons, L. (ed). (1991). *Assessing second language writing in academic context.* Norwood, NJ: Ablex Publishing Corp.

Hyland, K. (2003*). Second language writing*. London: Cambridge University Press.

Kroll, B. (ed.). (1990). *Second language writing: Research insights for the classroom*. Cambridge: Cambridge University Press.

McLaughlin, M. W. (1975). *Evaluation and reform: The elementary and secondary Education Act of 1965/Title I*. Cambridge, MA: Ballinger.

McNamara, T. F. (1996). *Measuring second language performance*. London: Addison Wesley Longman.

Peha, S. (2011). *Assessing writing, assessing writers*. Retrieved from http://www.ttms.org/PDFs/09%20Writing%20Assessment%20v001%20%28Full%29.pdf

Salmani Nodoushan, M. A. (2007a). Error treatment in the EFL writing class: Red pen method versus remedial instruction. *Journal of Educational Technology, 4*(3), 53-58.

Salmani Nodoushan, M. A. (2007b). Thinking on the write path. *Training Journal, May 2007*: 37-40.

Swales, J. M., & Feak, C. B. (1994). *Academic writing for graduate students: A course for non-native speakers of English*. Michigan: The University of Michigan Press.

Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.

White, E. (1994). *Teaching and assessing writing*. San Francisco: Jossey Bass.

THE AUTHOR

Dr. Mohammad Ali Salmani Nodoushan is an associate professor of Applied Linguistics at the Iranian Institute for Encyclopaedia Research, Tehran, Iran. His main areas of interest include politeness and pragmatics. He has published over 40 papers in international academic journals including Teaching and Teacher Education, Speech Communication, TESL Canada Journal, and so on. In addition, he has (co)authored three academic books and also authored one on IELTS. Besides his position as the editor in chief of the International Journal of Language Studies, he sits on the editorial boards of the Journal of Asia TEFL, Journal of Linguistic and Intercultural Education, Asian EFL Journal, and Journal on English Language Teaching. He is also a reviewer for a number of journals including the Journal of Pragmatics, TESOL Quarterly, Pragmatics and Society, and The Journal of Politeness Research.