

# Scoring Procedures for Assessing Writing

Taejoon Park

## Overview

The evaluation of writing ability of L2 students has become increasingly important in recent years, because the results of such evaluations are used for a variety of administrative, instructional, and research purposes. One of the first decisions to be made in determining a system for directly assessing writing quality is what type of scoring procedure will be used: Should a single score be given to each text, or should the different features of a text be scored separately? This issue has been the subject of a great deal of research and discussion in the composition literature. Among others, two types of scoring procedures have mainly been discussed: analytic and holistic scoring.

## Holistic Scoring

An efficient and increasingly popular approach to writing assessment is holistic scoring, which aims to rate the *overall proficiency level* reflected in a given sample of student writing. In a typical holistic scoring session, each writing sample is read quickly and then judged against a rating scale, or scoring rubric, that outlines the scoring criteria. Holistic scoring rubrics generally consist of 4 to 10 levels or bands, each of which corresponds to a score and a set of descriptors. These descriptors in the rubric can be either general or fairly specific. The existence of a scoring rubric distinguishes holistic scoring from its predecessor, *general impression marking*, in which criteria are never explicitly stated.

Holistic scoring is more economical than analytic scoring since readers are required to make only one decision (i.e., a single score) for each writing sample. For this reason, holistic scoring is commonly used in large-scale assessment of writing.

The major disadvantage of holistic scoring emerges from the limitations of the single score, which gives useful ranking information but no details. That is, holistic scoring cannot provide useful diagnostic information about a person's writing ability, as a single score does not allow raters to distinguish between various aspects of writing such as control of syntax, depth of vocabulary, organization, and so on. This is especially problematic for second-language writers since different aspects of writing ability may develop at different rates for different L2 learner. Hence, the same holistic score assigned to two different texts may represent two entirely different distinct sets of characteristics, even if the raters' scores reflect a strict and consistent application of the rubric.

Despite the drawbacks mentioned above, researchers in both L1 and L2 writing generally agree that holistic scoring is reliable, provided guidelines pertaining to rater training and rating session administration are faithfully adhered to (Perkins, 1983; White, 1994). Opinions on the validity of the procedure, however, tend to vary. Charney (1984), for example, states that "the validity of holistic scoring remains an open question" (p.67), while Perkins (1983) claims that

“holistic scoring has the highest construct validity when *overall attained writing proficiency* is the construct assessed” (p.652).

### **Analytic Scoring**

Analytic scoring procedures involve the separation of the various features of a composition into components for scoring purposes. Depending on the purpose of the assessment, texts might be rated on such features as content, organization, cohesion, register, vocabulary, grammar, or mechanics. Analytic scoring schemes thus provide more detailed information about a test taker’s performance in different aspects of writing.

Analytic scoring is preferred over holistic schemes by many writing specialists for a number of reasons. First, as mentioned above, it provides more useful diagnostic information about students’ writing abilities. That is, it provides more information about the strengths and weaknesses of students, and thus allow instructors and curriculum developers to tailor instruction more closely to the needs of their students. Second, analytic scoring is particularly useful for L2 learners, who are more likely to show a *marked* or *uneven profile* across different aspects of writing (e.g., some L2 learners may have excellent writing skills in terms of content and organization, but may have much lower grammatical control; others may have an excellent control of sentence structure, but may not know how to organize their writing in a logical way). Third, it is easier to train raters to use analytic scoring schemes, by virtue of such schemes’ explicit criteria in separate components, than to train raters to use holistic rubrics (Cohen 1994; McNamara, 1996). For example, inexperienced raters may find it easier to work with an analytic scale than a holistic rubric because they can evaluate specific textual features. Finally, the explicitness of analytic scoring guides offers teachers a potentially valuable tool for providing writers with consistent and direct feedback.

The major disadvantage of analytic scoring is that it takes longer than holistic scoring since readers are required to make more than one decision for every writing sample. Critics of analytic scoring also point out that measuring the quality of a text by tallying accumulated subskill scores diminishes the interconnectedness of written discourse, and gives the false impression that writing can be understood and fairly assessed by analyzing autonomous text features (Hillocks, 1995; White, 1994). Consequently, component scales may not be used effectively according to their internal criteria, resulting in a halo effect in which one component score may influence another. An additional problem with some analytic scoring schemes is that even experienced essay judges sometimes find it difficult to assign numerical scores based on certain descriptors (Hamp-Lyons, 1989). Thus, qualitative judgments about coherence, style, and so on are not always easily accommodated by analytic scoring methods.

### **Conclusion**

There are several scoring schemes currently in use for writing assessment. This article has focused on the two main types of scoring schemes: holistic and analytic scoring. It should be noted that no test or scoring procedure is suitable for all purposes, and even with guidelines and set criteria, the holistic and analytical scoring procedures can yield unreliable and invalid test

information. Therefore, decisions regarding the selection of evaluation procedures need to be made within the context of the specific testing situation.

## REFERENCES

- Charney, D. (1984). The validity of using holistic scoring to evaluate writing: A critical overview, *Research in the Teaching of English*, 18, 65-81.
- Cohen, A. (1994). *Assessing language ability in the classroom*. Boston: Heinle & Heinle.
- Hamp-Lyons, L. (1989). Raters respond to rhetoric in writing. In H.W. Dechert & Raupauch (Eds.), *Interlingual processes* (pp. 229-244). Tübingen: Gunter Narr.
- Hillocks, G., Jr. (1995). *Teaching writing as reflective practice*. New York: Teachers College Press.
- McNamara, T. (1996). *Measuring second language performance*. London: Longman.
- Perkins, K. (1983). On the use of composition scoring techniques, objective measures, and objective tests to evaluate ESL writing ability. *TESOL Quarterly*, 17, 651-671.
- White, E.M. (1994). *Teaching and assessing writing*. San Francisco, CA: Jossey-Bass.