Assessing Foreign/Second Language Writing Ability

Christine Coombe
Christine Coombe has a PhD in Foreign and Second Language Education from The Ohio State University, USA. She is a faculty member at Dubai Men’s College, Higher Colleges of Technology, UAE. She is the president-elect of TESOL. She has published in the areas of language assessment, teacher effectiveness, language teacher research, task-based learning and leadership in ELT.
Assessing writing skills is one of the most problematic areas in language testing. It is made even more important because good writing ability is very much sought after by higher education institutions and employers. To this end, good teachers spend a lot of time ensuring that their writing assessment practices are valid and reliable. This paper explores the main practical issues that teachers often face when evaluating the writing work of their students. It will consider issues and solutions in five major areas: test design; test administration; ways to assess writing; feedback to students; and effects on pedagogy.

Test Design

Approaches to Writing Assessment

The first step in test design is for teachers to identify which broad approach to writing best identifies their chosen type of assessment: direct or indirect. *Indirect writing assessment* measures correct usage in sentence level constructions and focuses on spelling and punctuation via objective formats like MCQs and cloze tests. These measures are supposed to determine a student’s knowledge of writing sub skills such as grammar and sentence construction which are assumed to constitute components of writing ability. Indirect writing assessment measures are largely concerned with accuracy rather than communication.

*Direct writing assessment* measures a student’s ability to communicate through the written mode based on the production of written texts. This type of writing assessment requires the student to come up with the content, find a way to organize the ideas, and use appropriate vocabulary, grammatical conventions and syntax. Direct writing assessment integrates all elements of writing. The choice of one approach over another should inform all subsequent choices in assessment design.

Aspects of Assessment Design

According to Hyland (2003), the design of good writing assessment tests and tasks involves four basic elements: rubric; prompt; expected response; and post-task evaluation. In addition, topic restriction should be considered here.

**Rubric.** The rubric is the instructions for carrying out the writing task. One problem for the test writer is to decide what should be covered in the rubric. A good rubric should include information such as the procedures for responding, the task format, time allotted for completion of the test/task, and information about how the test/task will be evaluated. Much of the information in the rubric should come from the test specification. Test specifications for a writing test should provide the test writer with details on the topic, the rhetorical pattern to be tested, the intended audience, how much information should be included in the rubric, the number of words the student is expected to produce, and overall weighting (Davidson & Lloyd, 2005). Good rubrics should:

- Specify a particular rhetorical pattern, length of writing desired and amount of time allowed to complete the task;
- Indicate the resources student will have available at their disposal (dictionaries, spell/grammar check etc) and what method of delivery the assessment will take (i.e. paper and pencil, laptop, PC);
- Indicate whether a draft or an outline is required;
- Include the overall weighting of the writing task as compared to other parts of the exam.
**Figure 1: Sample Writing Test Specification**

**Writing prompt.** Hyland (2003) defines the prompt as “the stimulus the student must respond to” (p. 221). Kroll and Reid (1994, p. 233) identify three main prompt formats: base, framed and text-based. The first two are the most common in F/SL writing assessment. Base prompts state the entire task in direct and very simple terms: for example, “Many say that “money is the root of all evil.” Do you agree or disagree with this statement?” Framed prompts present the writer with a situation that acts as a frame for the interpretation of the task: *On a recent flight back home to the UAE, the Airline lost your baggage. Write a complaint letter to Mr. Al-Ahli, the General Manager, telling him about your problem. Be sure to include the following:*… Text-based prompts present writers with a text to which they must respond or utilize in their writing. You have been put in charge of selecting an appropriate restaurant for your senior class party. Use the restaurant reviews below to select an appropriate venue and then write an invitation letter to your fellow classmates persuading them to join you there.

A writing prompt, irrespective of its format, defines the writing task for students. It consists of a question or a statement that students will address in their writing and the conditions under which they will be asked to write. Developing a good writing prompt requires an appropriate ‘signpost’ term, such as *describe, discuss, explain, compare, outline*, evaluate and so on to match the rhetorical pattern required. Each prompt should meet the following criteria:

- generate the desired type of writing, genre or rhetorical pattern;
- get students involved in thinking and problem-solving;
- be accessible, interesting and challenging to students;
- address topics that are meaningful, relevant, and motivating;
- not require specialist background knowledge;
- use appropriate signpost verbs;
- be fair and provide equal opportunities for all students to respond
- be clear, authentic, focused and unambiguous;
- specify an audience, a purpose, and a context.

(Davidson & Lloyd, 2005)

**Expected response.** This is a description of what the teacher intends students to do with the writing task. Before communicating information on the expected response to students, it is necessary for the teacher to have a clear picture of what type of response they want the assessment task to generate.

**Post task evaluation.** Finally, whatever way is chosen to assess writing, it is recommended that the effectiveness of the writing tasks/tests is evaluated. According to Hyland (2003), good writing tasks are likely to produce positive responses to the following questions:

- Did the prompt discriminate well among my students?
- Were the essays easy to read and evaluate?
- Were students able to write to their potential and show what they knew?
**Topic restriction.** This is in addition to the four aspects of test design described above. Topic restriction is a controversial and often heated issue in writing assessment. Topic restriction is the belief that all students should be asked to write on the same topic with no alternatives allowed. Many teachers may believe that students perform better when they have the opportunity to select the prompt from a variety of alternative topics. When given a choice, students often select the topic that interests them and one for which they have background knowledge. The obvious benefit of providing students with a list of alternatives is that if they do not understand a particular prompt, they will be able to select another. The major advantage to giving students a choice of writing prompt is the reduction of student anxiety.

On the other hand, the major disadvantage of providing more than one prompt is that it is often difficult to write prompts which are at the same level of difficulty. Many testers feel that it is generally advisable for all students to write on the same topic because allowing students to choose topics introduces too much variance into the scores. Moreover, marker consistency may be reduced if all papers read at a single writing calibration session are not on the same topic. It is the general consensus within the language testing community that all students should write on the topic and preferably on more than one topic. Research results, however, are mixed on whether students write better with single or with multiple prompts (Hamp-Lyons, 1990). It is thought that the performance of students who are given multiple prompts may be less than expected because students often waste time selecting a topic instead of spending that time writing. If it is decided to allow students to select a topic from a variety of alternatives, alternative topics should be of the same genre and rhetorical pattern. This practice will make it easier to achieve inter-rater reliability.

**Test Administration Conditions**

Two key aspects can be considered here: test delivery mode, including the resources made available to students; and time allocation.

**Test delivery mode.** In this day and age, technology has the potential to impact writing assessment. In the move toward more authentic writing assessment, it is being argued that students should be allowed to use computer, spell and grammar check, thesaurus, and online dictionaries as these tools would be available to them in real-life contexts. In parts of the world where writing assessment is taking place electronically, these technological advances bring several issues to the fore. First of all, when we allow students to use computers, they have access to tools such as spell and grammar check. This access could put those who write by hand at a distinct disadvantage. The issue of skill contamination must also be considered as electronic writing assessment is also a test of keyboarding and computer skills. Whatever delivery mode you decide to use for your writing assessments, it is important to be consistent with all students.

**Time allocation.** A commonly-asked question by teachers is how much time should students be given to complete writing tasks. Although timing would depend on whether you are assessing process or product, a good rule of thumb is provided by Jacobs et al (1981). In their research on the Michigan Composition Test, they (1981:19) state that allowing 30 minutes is probably sufficient time for most students to produce an adequate sample of writing. With process oriented writing or portfolios, much more time should be allocated for assessment tasks.

**Ways to Assess Writing**

The assessment of writing can range from the personalized, holistic and developmental on the one hand to carefully quantified and summative on the other hand. In the following section, the assessment benefits of student-teacher conferences, self-assessment, peer assessment, and portfolio assessment are considered, each of these verging towards the holistic end of the
assessment cline. In addition, however, the role of rating scales and depersonalized or objective marking procedures needs to be considered.

**Student-teacher conferences.** Teachers can learn a lot about their students’ writing habits through student-teacher conferences. These conferences can also provide important assessment opportunities. Among the questions that teachers might ask during conferences include:

- How did you select this topic?
- What did you do to generate content for this writing?
- Before you started writing, did you make a plan or an outline?
- During the editing phase, what types of errors did you find in your writing?
- What do you feel are your strengths in writing?
- What do you find difficult in writing?
- What would you like to improve about your writing?

**Self assessment.** There are two self-assessment techniques than can be used in writing assessment: dialog journals and learning logs. Dialog journals require students to regularly make entries addressed to the teacher on topics of their choice. The teacher then writes back, modeling appropriate language use but not correcting the student’s language. Dialog journals can be in a paper/pencil or electronic format. Students typically write in class for a five to ten minute period either at the beginning or end of the class. If you want to use dialog journals in your classes, make sure you don’t assess students on language accuracy. Instead, Peyton and Reed (1990) recommend that you assess students on areas like topic initiation, elaboration, variety, and use of different genres, expressions of interests and attitudes, and awareness about the writing process.

**Peer assessment.** Peer Assessment is yet another technique that can be used when assessing writing. Peer assessment involves the students in the evaluation of writing. One of the advantages of peer assessment is that it eases the marking burden on the teacher. Teachers don’t need to mark every single piece of student writing, but it is important that students get regular feedback on what they produce. Students can use checklists, scoring rubrics or simple questions for peer assessment. The major rationale for peer assessment is that when students learn to evaluate the work of their peers, they are extending their own learning opportunities.

**Portfolio assessment.** As far as portfolios are defined in writing assessment, a portfolio is a purposive collection of student writing over time, which shows the stages in the writing process a text has gone through and thus the stages of the writers’ growth.

Several well-known testers have put forth lists of characteristics that exemplify good portfolios. For instance, Paulson, Paulson and Meyer (1991) believe that portfolios must include student participation in four important areas: 1) the selection of portfolio contents; 2) the guidelines for selection; 3) the criteria for judging merit and 4) evidence of student reflection. The element of reflection figures prominently in the portfolio assessment experience. By having reflection as part of the portfolio process, students are asked to think about their needs, goals, weaknesses and strengths in language learning. They are also asked to select their best work and to explain why that particular work was beneficial to them. Learner reflection allows students to contribute their own insights about their learning to the assessment process. Perhaps Santos (1997) says it best, “Without reflection, the portfolio remains ‘a folder of all my papers’ (p. 10).

**Marking Procedures for Formal and Summative Assessment**

The following section considers who will mark formal assessment of the student, the types of scales that can be established for assessment, and the procedures for assessing the scripts.
Classroom teacher as rater. While students, peers and the learner herself can legitimately be included in assessment designed primarily for developmental purposes, such as described above, different issues arise when it comes to formal and summative assessment. Should classroom teachers mark their own students’ papers? Experts disagree here. Those who are against having teachers mark their own students’ papers warn that there is the possibility that teachers might show bias either for or against a particular student. Other experts believe that it is the classroom teacher who knows the student best and should be included as a marker. Double blind marking is the recommended ideal where no student identifying information appears on the scripts.

Multiple raters. Do we really need more than one marker for student writing samples? The answer is an unequivocal ‘yes’. All reputable writing assessment programs use more than one rater to judge essays. In fact, the recommended number is two, with a third in case of extreme disagreement or discrepancy. Why? It is believed that multiple judgments lead to a final score that is closer to a “true” score than any single judgment (Hamp-Lyons, 1990).

Establish Assessment Scales. An important part of writing assessment deals with selecting the appropriate writing scale. Selecting the appropriate marking scale depends upon the context in which a teacher works. This includes the availability of resources, amount of time allocated to getting reliable writing marks to administration, and the teacher population and management structure of the institution. The F/SL assessment literature generally recognizes two different types of writing scales for assessing student written proficiency: holistic and analytic.

Holistic Marking Scales. Holistic marking is based on the marker’s total impression of the essay as a whole. Holistic marking is variously termed as impressionistic, global or integrative marking. Experts in holistic marking scales recommend that this type of marking is quick and reliable if 3 to 4 people mark each script. The general rule of thumb for holistic marking is to mark for two hours and then take a rest grading no more than 20 scripts per hour. Holistic marking is most successful using scales of a limited range (e.g., from 0-6).

F/SL educators have identified a number of advantages to this type of marking. First, it is reliable if done under no time constraints and if teachers receive adequate training. Also, this type of marking is generally perceived to be quicker than other types of writing assessment and enables a large number of scripts to be scored in a short period of time. Third, since overall writing ability is assessed, students are not disadvantaged by one lower component such as poor grammar bringing down a score. An additional advantage is that the scores tend to emphasize the writer’s strengths (Cohen, 1994: 315).

Several disadvantages of holistic marking have also been identified. First of all, this type of marking can be unreliable if marking is done under short time constraints and with inexperienced, untrained teachers (Heaton, 1990). Secondly, Cohen (1994) has cautioned that longer essays often tend to receive higher marks. Testers point out that reducing a score to one figure tends to reduce the reliability of the overall mark. It is also difficult to interpret a composite score from a holistic mark. The most serious problem associated with holistic marking is the inability of this type of marking to provide washback. More specifically, when marks are gathered through a holistic marking scale, no diagnostic information on how those marks were awarded appears. Thus, testers often find it difficult to justify the rationale for the mark. Hamp-Lyons (1990) has stated that holistic marking is severely limited in that it does not provide a profile of the student’s writing ability. Finally, since this type of scale looks at writing as a whole, there is the tendency on the part of the marker to overlook the various sub-skills that make up writing. For further discussion of these issues, see both Educational Testing Service (ETS) and the International English Language Testing Systems (IELTS) research publications in the area of holistic marking.

Analytical Marking Scales. Analytic marking is where “raters provide separate assessments for each of a number of aspects of performance” (Hamp-Lyons, 1991). In other words, raters mark selected aspects of a piece of writing and assign point values to quantifiable criteria. In
the literature, analytic marking has been termed discrete point marking and focused holistic marking. Analytic marking scales are generally more effective with inexperienced teachers. In addition, these scales are more reliable for scales with a larger point range.

A number of advantages have been identified with analytic marking. Firstly, unlike holistic marking, analytical writing scales provide teachers with a “profile” of their students’ strengths and weaknesses in the area of writing. Additionally, this type of marking is very reliable if done with a population of inexperienced teachers who have had little training and grade under short time constraints (Heaton, 1990). Finally, training raters is easier because the scales are more explicit and detailed.

Just as there are advantages to analytic marking, educators point out a number of disadvantages associated with using this type of scale. Analytic marking is perceived to be more time consuming because it requires teachers to rate various aspects of a student’s essay. It also necessitates a set of specific criteria to be written and for markers to be trained and attend frequent calibration sessions. These sessions are to insure that inter-marker differences are reduced which thereby increase validity. Also, because teachers look at specific areas in a given essay, the most common being content, organization, grammar, mechanics and vocabulary, marks are often lower than for their holistically-marked counterparts.

Perhaps the most well-known analytic writing scale is the ESL Composition Profile (Jacobs et al, 1981). This scale contains five component skills, each focusing on an important aspect of composition and weighted according to its approximate importance: content (30 points), organization (20 points), vocabulary (20 points), language use (25 points) and mechanics (5 points). The total weight for each component is further broken down into numerical ranges that correspond to four levels from “very poor” to “very good” to “excellent”.

*Establish procedures and a rating process.* Reliable writing assessment requires a carefully thought-out set of procedures and a significant amount of time needs to be devoted to the rating process.

First, a small team of trained and experienced raters needs to select a number of sample benchmark scripts from completed exam papers. These benchmark scripts need to be representative of the following levels at minimum:

- **Clear pass** (good piece of writing that is solidly in the A/B range)
- **Borderline pass** (a paper that is on the borderline between pass and fail but shows enough of the requisite information to be a pass)
- **Borderline fail** (a paper that is on the borderline between pass and fail but does not have enough of the requisite information to pass)
- **Clear fail** (a below average paper that is clearly in the D/F range)

Once benchmark papers have been selected, the team of experienced raters needs to rate the scripts using the scoring criteria and agree on a score. It will be helpful to note down a few of the reasons why the script was rated in such a way. Next, the lead arbitrator needs to conduct a calibration session (oftentimes referred to as a standardization or norming session) where the entire pool of raters rate the sample scripts and try to agree on the scores that each script should receive. In these calibration sessions, teachers should evaluate and discuss benchmark scripts until they arrive at a consensus score. These calibration sessions are time consuming and not very popular with groups of teachers who often want to get started on the writing marking right away. They can also get very heated especially when raters of different educational and cultural backgrounds are involved. Despite these disadvantages, they are an essential component to standardizing writing scores.
Responding to Student Writing

Another important aspect of writing marking is providing written feedback to students. This feedback is essential in that it provides opportunities for students to learn and make improvements to their writing. Probably the most common type of written teacher feedback is handwritten comments on the students’ papers. These comments usually occur at the end of the paper or in the margins. Some teachers like to use correction codes to provide formative feedback to students. These simple correction codes facilitate marking and minimize the amount of ‘red ink’ on student writings.

Figure 2 is an example of a common correction code used by teachers. Advances in technology provide us with another way of responding to student writing. Electronic feedback is particularly valuable because it can be used to give a combination of handwritten comments and correction codes. Teachers can easily provide commentary and insert corrections through Microsoft Word’s track changes facility and through simple-to-use software programs like Markin (2002).

<table>
<thead>
<tr>
<th>sp</th>
<th>Spelling</th>
</tr>
</thead>
<tbody>
<tr>
<td>vt</td>
<td>Verb tense</td>
</tr>
<tr>
<td>ww</td>
<td>Wrong word</td>
</tr>
<tr>
<td>wv</td>
<td>Wrong verb</td>
</tr>
<tr>
<td>😊</td>
<td>Nice idea/content!</td>
</tr>
<tr>
<td>☐</td>
<td>Switch placement</td>
</tr>
<tr>
<td>¶</td>
<td>New paragraph</td>
</tr>
<tr>
<td>?</td>
<td>I don’t understand</td>
</tr>
</tbody>
</table>

Figure 2: Sample Marking Codes for Writing

Research indicates that teacher written feedback is highly valued by second language writers (Hyland, 1998 as cited in Hyland, 2003) and many students particularly value feedback on their grammar (Leki, 1990). Although positive remarks are motivating and highly valued by students, Hyland (2003) points out that too much praise or positive commentary early on in a writer’s development can make students complacent and discourage revision (p. 187).

Effects on Pedagogy

The cyclical relationship between teaching and assessment can be made entirely positive provided that the assessment is based on sound principles and procedures. Both teaching and assessment should relate to the learners’ goals and very frequently to institutional goals.

**Process Versus Product.** The goals of all the stakeholders can be met when a judicious balance is established, in the local context, between process and product. In recent years, there has been a shift towards focusing on the process of writing rather than on the written product. Some writing tests have focused on assessing the whole writing process from brainstorming activities all the way to the final draft (or finished product). In using this process approach, students usually have to submit their work in a portfolio that includes all draft material. A more traditional way to assess writing is through a product approach. This is most frequently accomplished through a timed essay, which usually occurs at the mid and end point of the semester. In general, it is recommended that teachers use a combination of the two approaches in their teaching and assessment, but the approach ultimately depends on the course objectives.

**Some Aspects of Good Teacher-Tester Practice.** Teachers and testers know that any type of assessment should first and foremost reflect the goals of the course, so they start the test
development process by reviewing their test specifications. They will avoid a “snap shot” approach to writing ability by giving students plenty of opportunities to practice a variety of different writing skills. They will practice multiple-measures writing assessment by using tasks which focus on product (e.g., essays at midterm and final) and process (e.g., writing portfolio). They will give more frequent writing assessments because they know that assessment is more reliable when there are more samples to assess. They will provide opportunities for a variety of feedback from teacher and peers, as well as opportunities for the learners to reflect on their own progress. Overall, they will ensure that the learner’s focus is maintained primarily on the learning process and enable them to see that the value of the testing process is primarily to enhance learning by measuring real progress and identifying areas where further learning is required.

References


