

Ms. Wright's Comments on Mr. Knott's Reading Test

First, while I applaud Mr. Knott's attempt to make his students more aware of current events, he needs to be more selective about the topic. A text about alligator attacks is certain to raise some anxiety among students in Florida! No doubt Mr. Knott intended to provide a public safety message, but this text would have been more appropriate for a class discussion where students could be reassured about the actual dangers.

That said, I'd like to commend Mr. Knott for attempting to assess a wide range of reading subskills and not just focus on specific details, which is so often the case. He evenly uses most sections of the text and asks questions about both grammar and vocabulary in context. I appreciate his focus on critical thinking, both in the labeled section and in other questions based on inferring skills. However, I wonder if it is advisable to label these as something special? Ideally, all questions would utilize critical thinking.

My main concerns have to do with test and item construction. Although Mr. Knott did follow the specifications with regard to subskills to be covered, the organization of his test is quite chaotic. In a reading or listening exam, it is important to stick to one format for subsections of the exam so that students don't constantly have to adjust to task changes. Within each subsection, the questions should occur in the same order as the content of the passage. Next time, Mr. Knott should mark up the text so he can clearly see the progression of testable points.

There are also technical problems with his test. There is no clear indication of how many points each question or section is worth. Some sections have instructions while others don't. Even those with instructions have problems. For example, what does Mr. Knott intend by "DK—don't know"? Perhaps he has confused it with "NG," meaning that information is not given in the passage. On the other hand, maybe he means that the question is unanswerable by the student. Item 4 in that section is especially problematic because of the implied double negative with "cannot" and a negative response. Mr. Knott could also be more consistent with the formatting of his MCQs.

In several places there are problems with keys, the intended answers. In question 2, either 17 or 20 would be correct since he doesn't specify the time frame. In question 7, all of the answers are plausible since it is not clear which they he intends to reference in paragraph 3. Paraphrasing is an important reading skill, and it is good that Mr. Knott chose to test it. However, in item 2, will students understand the meaning of "fatal" in the text and relate it to "died"? In item 8, will they understand "least"? In addition, some distractors could be eliminated without students even reading the text. Lastly, Mr. Knott forgot to credit his sources.

4 Assessing Writing

How important is writing in your school's assessment of your learners?

(A) Intensive English Program, Community College, University	(B) Conversation Classes, Non-academic Programs, Survival Classes: Writing is not important and is rarely tested. However, learners may need a certain level of writing ability for certain short answer or short essay questions.	(C) K-12: Along with reading, writing is considered one of the most important academic skills that all students are expected to master. Test scorers need to realize that a certain number of language errors will persist for some time with ELLs, namely errors with articles, prepositions, and verb tenses. These errors are peculiar to ELLs and should not be graded as harshly as material that was taught to everyone such as spelling or punctuation. Teachers should be aware of the importance of a clear, easy-to-understand, culturally-neutral prompt for the exam. Content area teachers from science and social studies should realize that any essay question will almost always produce a lower score from ELLs because of their language problems even when they know the content material well.
Writing is fundamental to academic success and is certainly a major component of any type of assessment. Many post-secondary programs require a writing sample before admission. The TOEFL requires several writing samples from students.		

The field of writing assessment has developed considerably over the past 50 years. Teachers ask their students to write on a variety of topics and then assess them on the message contained in the writing sample, the clarity and organization of the message, and the mechanics (spelling, capitalization, and punctuation) utilized (O'Malley & Valdez Pierce, 1996). This chapter explores the practical issues that teachers face when evaluating the writing work of their students.

Assessing writing skills is important because good writing ability is highly sought by higher education institutions and employers. To this end, Ms. Wright spends a lot of time ensuring that her writing assessment practices are valid and reliable. She knows that any type of assessment should reflect the course goals, so she starts the test development process by reviewing the course outcomes and the specifications. Some of the things she does to ensure valid and reliable writing assessment are:

- She avoids a "snapshot" approach of assessing writing ability by giving students plenty of opportunities to practice a variety of writing skills.
- She practices multiple-measures writing assessment by using tasks that focus both on product (midterm and final essays) and process (writing portfolio).
- She gives frequent writing assessments because she knows that assessment is more reliable when there are more samples to assess.
- She doesn't give students a choice of writing topic on important exams like midterm and final because the marking will be more difficult to standardize if she does.
- She avoids using a red pen to mark her students' papers as she feels it does more harm than good. When papers are submitted electronically, she uses computerized marking programs.
- Before marking begins, she identifies benchmark papers and later shows them to students during instruction.
- She uses a well-established analytic marking scale at midterm because several inexperienced teachers will be helping grade papers.
- She makes sure that at least two teachers mark every writing test.

Approaches to Writing Assessment

Two major approaches to writing assessment have been identified in the literature—indirect and direct.

Indirect measures of writing assessment assess correct usage in sentence-level constructions and assess spelling and punctuation via objective formats like multiple choice 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 more concerned with accuracy than communication. In the past, tests such as TOEFL[®] used indirect writing assessment where four sections of a sentence were underlined and marked as "A," "B," "C," and "D." One of these contained an error, and students had to identify the error.

Direct measures of writing assessment assess a student's ability to communicate through the written mode based on the actual production of written texts. This type of writing assessment requires the student to produce 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.

Considerations in Designing Writing Assessment Tasks

According to Hyland (2003), the design of good writing assessment tests and tasks involves four basic elements:

- *Rubric*: the instructions
- *Prompt*: the task
- *Expected response*: what the teacher intends students to do with the task
- *Post-task evaluation*: assessing the effectiveness of the writing task.

The first element of a good writing assessment is the *rubric*, the instructions for carrying out the writing task. A rubric can also mean the set of criteria on which a piece of work, such as a project, is evaluated, and it is used in this sense in elementary education. Good writing instructions or rubrics should:

- specify a particular rhetorical pattern, length of writing desired, and amount of time allowed to complete the task
- indicate the resources students will have available at their disposal (dictionaries, spell/grammar checker, etc.), and the delivery method of the assessment (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

Most 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 range of topics, 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). Here is an example related to an essay on travel.

Sample Writing Test Specification

Topic	Related to the theme of travel
Text Type	Compare/contrast
Length	250 words
Areas to be assessed	Content, organization, vocabulary, language use, mechanics
Timing	30 minutes
Weighting	10 percent of midterm exam grade
Pass level	Similar to IELTS™ band 5.5

The second essential part of any test of writing is the *writing prompt*. Hyland (2003) defines the prompt as “the stimulus the student must respond to” (p. 221). Kroll and Reid (1994) identify three main prompt formats: base, framed, and text-based (p. 233). The first two are the most common in second/foreign language writing assessment. Base prompts state the entire task in direct and simple terms, whereas framed prompts present the writer with a situation that acts as a frame for the interpretation of the task. Text-based prompts present writers with a text to which they must respond or utilize in their writing. Consider these examples.

Base Prompts

- Do you favor or oppose a complete ban on smoking? Why? Why not?
- Discuss the view that women are better drivers than men.
- Many say that “Money is the root of all evil.” Do you agree or disagree with this statement?

Framed Prompts

- On a recent flight back home to the UAE, Emirates Airlines 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:
 - your flight details
 - a description of the baggage lost and its contents
 - what you would like Mr. Al-Ahli to do for you
- An announcement has been made by the Academy on this year’s nominations for the Oscar awards. You and most of your friends have not heard of the films nominated for the Best Picture award. Write a letter to the head of the Academy Awards about this, and suggest and support alternative nominees.

Text-Based Prompts

- You have been asked by a youth travel magazine to write an article about things to see and do in your hometown. Using the attached set of pictures, write a one-page article on this topic.
- You have been put in charge of selecting an appropriate restaurant for your senior class party. Use the restaurant reviews to select an appropriate venue, and then write an invitation letter to your fellow classmates persuading them to join you there.

Criteria of Good Writing Prompts

Each prompt you use in the assessment of writing 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

Source: Davidson, P., & Lloyd, D. (2005). Guidelines for developing a reading test. In D. Lloyd, P. Davidson, & C. Coombe (Eds.), *The fundamentals of language assessment: A practical guide for teachers in the Gulf* (pp. 53–63). Dubai: TESOL Arabia Publications.

Developing a good writing prompt requires that you use the appropriate signpost term to match the rhetorical pattern you are using. Some of the most common signpost terms are:

- Describe:** give a detailed account
- Discuss:** argue a thesis, identifying pros and cons
- Explain:** state and interpret
- Compare:** show similarities between two things
- Contrast:** show the differences between two things
- Analyze:** identify main points and evaluate them
- Define:** provide the definition and exemplify
- Summarize:** produce a concise account of the main ideas, omitting details and examples
- Outline:** provide a summary of main points and sub-headings
- Evaluate:** appraise the worth or value of something

A third essential element of good writing assessment is the *expected response*, a description of what the teacher intends students to do with the writing task. Before communicating information on the expected response to students, the teacher must have a clear picture of the type of response the assessment task should generate.

Finally, whatever way you choose to assess writing, *evaluate the effectiveness* of your writing tasks/tests. 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?

Issues in Writing Assessment

Time Allocation

Teachers often ask how much time students should 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, Zinkgraf, Wormuth, Hartfiel, & Hughey, et al. (1981). In their research on the Michigan Composition Test, they state that allowing 30 minutes is probably sufficient time for most students to produce an adequate sample of writing (p. 19). With process-oriented writing or portfolios, much more time should be allocated for assessment tasks.

Process versus Product

In recent years, there has been a shift toward the process of writing rather than on the written product. Some writing tests assess the whole writing process, from brainstorming activities 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 usually accomplished through a timed essay at the middle and end-point of the semester. It is recommended that teachers use a combination of the two approaches in their writing assessment, but the approach ultimately depends on the course objectives.

Use of Technology

Technology has the potential to affect writing assessment. Students writing on computers regularly use spell and grammar checker, a thesaurus, or online dictionaries as tools.

As writing assessment increasingly takes place electronically, access to these tools becomes an issue. If students using computers have access to spell

and grammar checkers, does this 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 use for your writing assessments, it is important to be consistent with all students.

Topic Restriction

Topic restriction, the belief that all students should be asked to write on the same topic with no alternatives allowed, is a controversial issue in writing assessment. Some teachers believe that students perform better when they have the opportunity to select the prompt from a variety of alternatives. When given a choice, students often select a 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, thus reducing student anxiety.

On the other hand, the major disadvantage of providing more than one prompt is that it is difficult to write prompts that are at the same level of difficulty, thus creating variance in 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 consensus within the language testing community that all students should write on the same topic and preferably on more than one topic. However, research results are mixed on whether students write better with single or multiple prompts (Hamp-Lyons, 1990b). It is thought that the performance of students who are given multiple prompts may be lower than expected because students often waste time selecting a topic instead of writing. If you decide to allow students to select a topic from a variety of alternatives, make sure your alternative topics are the same genre and rhetorical pattern. This practice will make it easier for you to achieve inter-rater reliability.

Techniques for Assessing Writing

The ESL/EFL literature generally addresses two types of writing—free writing and guided writing. Free writing requires students to read a prompt that poses a situation and write a planned response based on a combination of background knowledge and knowledge learned from the course. Guided writing, in contrast, requires students to manipulate content that is provided in the prompt, usually in the form of a chart or diagram.

Guided Writing

Guided writing is a bridge between objective and subjective formats. This task requires teachers to be clear about what they expect students to do. Decide in advance whether mechanical issues such as spelling, punctuation, and capitalization matter when the task focuses on comprehension. Some important points to keep in mind for guided writing are:

- Be clear about the expected form and length of response (e.g., one paragraph, a 250-word essay, a letter).
- If you want particular information included, clearly specify it in the prompt (e.g., three causes and effects, two supporting details).
- Similarly, specify the discourse pattern(s) the students are expected to use (e.g., compare and contrast, cause and effect, description).
- Since guided writing depends on the students' manipulation of the information provided, ask them to provide something beyond the prompt such as an opinion, an inference, or a prediction.
- Be amenable to revising the anticipated answer even as you grade.

Free-Writing

All of these suggestions are particularly germane to free-writing. The goal for teachers is to elicit comparable products from students of different ability levels. Some important points to keep in mind for free-writing are:

- The use of multiple raters is especially important in evaluating free-writing. Agree on grading criteria in advance and calibrate before the actual grading session.
- Decide whether to use holistic, analytical, or a combination of the two as a rating scale for marking.
- If using a band or rating scale (see pages 82–83 for an example), adjust it to the task.
- Acquaint students with the marking scheme in advance by using it for teaching, grading homework and providing feedback. (**Remember:** in all cases, good assessment mirrors actual classroom instruction.)

- Teach good writing strategies by providing students with enough space for an outline, a draft, and the finished product.
- In ESL/EFL classrooms, be aware of cultural differences and sensitivities. Avoid contentious issues that might offend or disadvantage students.

Authentic Writing Assessment

Student-Teacher Conferences

Teachers can learn a lot about their students' writing habits through student-teacher conferences, which can also provide important assessment opportunities. The questions 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

Two self-assessment techniques can be used in writing assessment—dialogue journals and learning logs. *Dialogue journals* require students to regularly make entries addressed to the teacher on topics of their choice. The teacher writes back, modeling appropriate language use but not correcting the students' language. Dialogue journals can be in a paper/pencil or electronic format. Students typically write in class for a five- to ten-minute period at the beginning or end of the class. If you want to use dialogue journals in your classes, don't assess students on language accuracy. Instead, Peyton and Reed (1990) recommend that you assess students on areas like topic initiation, elaboration, variety, use of different genres, expressions of interests and attitudes, and awareness about the writing process. *Learning logs* document time students spend on various writing activities.

Peer Assessment

Peer assessment, yet another assessment technique, involves the students in the evaluation of writing. One advantage of peer assessment is that it eases the marking burden on the teacher. Teachers do not need to mark every single piece of student writing, but it is important that students receive 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-Based Assessment

Portfolio-based assessment examines multiple pieces of writing produced over time under different constraints rather than a single essay written in a specified time period. For assessment purposes, a portfolio is a collection of student writing over time that shows the stages in the writing process a text has gone through and thus the stages of the writers' growth. Portfolios reflect accomplishment relative to specific instructional goals or objectives. Portfolios can showcase a student's best work or display a collection of both drafts and final products to demonstrate progress and continued improvement.

Characteristics of a Portfolio

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.

By including 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 explain why it was beneficial to them. Learner reflection allows students to contribute their own insights about their learning to the assessment process. Santos (1997) says it best: "without reflection, the portfolio remains 'a folder of all my papers.'"

Marking Procedures for the Assessment of Writing

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 or papers from completed exam papers. These benchmark papers need to be representative of the following levels at a 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, a team of experienced raters marks the papers using the scoring criteria and agrees on a score. It will be helpful to note a few of the reasons why the paper was rated as it was. Next, the lead arbitrator conducts a calibration session (often referred to as a standardization or norming session) where the entire pool of raters rate the sample papers and try to agree on the scores that each paper should receive. In calibration sessions, teachers should evaluate and discuss benchmark papers until they arrive at a consensus score. These sessions are time consuming and not popular with teachers who want to get started marking right away. Sessions can also get heated, especially when raters of different educational and cultural backgrounds are involved. Despite these disadvantages, they are an essential component to standardizing writing scores.

Writing Assessment Scales

An important part of writing assessment deals with selecting the appropriate writing scale for a particular teaching context. Factors to consider include the availability of resources, amount of time allocated to getting reliable writing

marks to administration, the teacher population, and the management structure of the institution.

The two main types of writing scales for assessing student written proficiency are 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 believe that this type of marking is quick and reliable if three to four people mark each paper. The rule of thumb for holistic marking is to mark for two hours and then take a rest, grading no more than 20 papers per hour. Holistic marking is most successful using scales of a limited range (e.g., from 0–6).

Both the Educational Testing Service (ETS) and the International English Language Testing Systems (IELTS™) have conducted a tremendous amount of research in the area of holistic marking, and second and foreign language educators have identified a number of advantages to it. First, it is reliable if done under no time constraints and if teachers receive adequate training. Second, 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. Finally, the scores tend to emphasize the writer's strengths (Cohen, 1994, p. 315).

Several disadvantages of holistic marking have also been identified. First, it can be unreliable if marking is done under short time constraints and with inexperienced, untrained teachers (Heaton, 1990). Second, Cohen (1994) has cautioned that longer essays often tend to receive higher marks. Third, 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 its inability to provide washback. Specifically, when marks are assigned through a holistic marking scale, there is no diagnostic information on how they were awarded. Thus, testers often find it difficult to justify the rationale for the mark. Hamp-Lyons (1990a) 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.

See Figure 2 on page 82 for an example of a holistic marking scale produced by the National Admissions and Placement Office in the United Arab Emirates for the Common Educational Proficiency Assessment (CEPA).

Figure 2. CEPA Writing Descriptors

6	<ul style="list-style-type: none"> • Overall meaning of complex communication adequately conveyed. • Main and subsidiary points are clear and well organized, but may contain minor irrelevancies or inappropriacies. • A range of cohesive devices used, though not always accurately. • Generally accurate use of sentence structure, though range of complex sentences is limited. • Vocabulary choice generally adequate, but may be inadequate to express a wide range of ideas with precision. • Most of the time, appropriate choice of words, idioms and register gives the text a feeling of fluency. • Occasional errors in spelling may still occur. • Uses capital letters, periods, commas, apostrophes, parentheses, and bullets, with only occasional unobtrusive errors.
5	<ul style="list-style-type: none"> • Overall meaning of simple and more complex communication adequately conveyed, though clarity will vary. • Organization of text contributes to overall clarity. • Range of cohesive devices is attempted. • Simple sentences are generally correct; some complex sentences may be used, but not often accurately. • Errors in subject-verb agreement may still occur. • Vocabulary generally appropriate. • Appropriate choice of words, idioms and register occasionally gives glimpses of fluency. • Spelling errors still intrude but do not impair meaning. • Uses capital letters, periods, commas, and apostrophes appropriately, with only occasional errors.
4	<ul style="list-style-type: none"> • Meaning clear in straightforward communications; where content is more complex, meaning comes through only intermittently. • Simple cohesive devices used appropriately. • Can construct simple sentences but errors in subject-verb agreement and word order are frequent. • Appropriate choice of basic tenses. • Range of vocabulary becomes wider, but may be inappropriate. • Text is slanted. • Spelling errors intrude though words are mainly recognizable with effort. • Uses capital letters and periods almost without error; commas and apostrophes missing or misused.
3	<ul style="list-style-type: none"> • Meaning only clear in short, simple communications; becomes unclear if content becomes more complex. • Little or no evidence of cohesive devices. • Attempts simple sentences with some awareness of, but limited control of, basic sentence structure and word order. • Vocabulary limited to simplest personal or work-related topics. • Spelling of familiar words is generally accurate, but unfamiliar words may be unrecognizable. • Uses capital letters and periods most of the time.

2	<ul style="list-style-type: none"> • Can convey only the simplest ideas. • Begins to produce a few short sentences and phrases independently, but with little or no control of sentence structure. • Vocabulary limited to common words. • Can spell a few common words accurately. • Some evidence of punctuation, but usually inaccurate. • Can copy sentences accurately. • Can write legibly.
1	<ul style="list-style-type: none"> • Can convey information by positioning words correctly (e.g., name, title). • Essentially unable to make sentences or multi-word messages. • Can write words from memory in a limited range (e.g., name, address, job). • Can form letters accurately and independently, but confuses upper and lower case except when copying. • Can copy words accurately but cannot spell words not given. • Text is so short that only evidence of letter formation or word copying can be assessed.
0	<ul style="list-style-type: none"> • Any of the following • No sample available. • Whole text appears to be copied or memorized and bears no relation to the topic. • Illegible.

Analytical Marking Scales

In analytic marking, “raters provide separate assessments for each of a number of aspects of performance” (Hamp-Lyons, 1991). Scorers mark selected aspects of a piece of writing and assign point values to quantifiable criteria. Analytic marking scales are generally more effective with inexperienced teachers and more reliable for scales with a larger point range.

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

Just as there are advantages to analytic marking, educators point out a number of disadvantages associated with it. 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 scorers to be trained in frequent calibration sessions to ensure that inter-marker differences are reduced to increase validity. Also, teachers tend to focus on spe-

cific areas in an essay such as content, organization, grammar, mechanics, and vocabulary. Consequently, analytic marks are often lower than holistically marked papers.

The best-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 into numerical ranges that correspond to four levels from very poor to very good to excellent.

Classroom Teacher as Rater

Should classroom teachers mark their own students' papers? Experts disagree here. Those who are against this warn there is the possibility that teachers might show bias for or against a particular student. Other experts believe the classroom teacher knows the student best and should be included as a marker. Double-blind marking is the recommended ideal with no identifying student information 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. It is believed that multiple judgments lead to a final score that is closer to a "true" score than any single judgment (Hamp-Lyons, 1990b, p. 79).

Responding to Student Writing

Another essential aspect of marking is providing written feedback to students so they can learn and make improvements to their writing. Probably the most common type of written teacher feedback is handwritten comments at the end or in the margins of the students' papers. Some teachers like to use simple correction codes to provide formative feedback. These codes facilitate marking and

minimize the amount of "red" ink on student writings. An example of a common correction code used by teachers follows. Advances in technology provide us with another way of responding to student writing. Electronic feedback is particularly valuable because it can give a combination of handwritten comments and correction codes. Teachers can easily provide commentary and insert corrections through Microsoft Word's track changes function and through simple-to-use software programs like Markin (www.cict.co.uk/software/markin/).

Sample Marking Codes for Writing

sp	Spelling
vt	Verb tense
vw	Wrong word
wv	Wrong verb
☺	Nice idea/content!
⌈⌋	Switch placement
¶	New paragraph
?	I don't understand

Research indicates that teacher written feedback is highly valued by second language writers (F. Hyland, 1998, as cited in Hyland, 2003; Ferris, 2002, p. 135), 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).

Ten Things to Remember about Writing Assessment

- 1. Give students multiple writing assessment opportunities.**
Provide plenty of opportunities for students at all levels to practice the type of writing that you expect them to do on the writing test. Often teachers avoid writing until exam time because it is a lot of work to individually mark several drafts of the same essay.
- 2. Test a variety of writing skills and create tasks of varying lengths.**
Take more than one sample of students' writing. This reduces the variation in performance that might occur from task to task. It is widely believed that the performance on one task is not representative of students' overall writing ability. The more samples of writing ability on the test, the more reliable the test.
- 3. Develop prompts that are appropriate for the students.**
The prompts you select or develop should invite the desired type of writing. They should be realistic and sensitive to the cultural background of the student. Choose subjects within the realm of your students' experience.
- 4. Evaluate all answers to one question before going on to the next.**
This practice prevents a shifting of standards from one question to the next and helps the rater mark more consistently.
- 5. Mark only what the student has written.**
Don't be influenced by other factors in addition to the quality of the work, such as the quality or legibility of the handwriting.
- 6. Have a systematic approach for dealing with marking discrepancies.**
One such approach might be to take the average of the two raters for a small discrepancy and utilize a third rater if there is a big discrepancy.
- 7. Get students involved.**
Get students involved in developing and marking their writing tests. Have them suggest prompts they'd like to write on and get them involved in peer assessment. Share whatever scoring criteria and rubrics you use with students. Transparency can help students internalize the rubric so that it becomes a natural part of their editing process.
- 8. Provide students with diagnostic feedback.**
Use writing assessment results to identify what students can and cannot do well and provide this information to students. With analytic marking scales, you will have access to a profile to give students feedback. With holistic marking scales, be sure to take notes on students' strengths and areas for improvement.
- 9. Practice blind or double blind marking.**
Mark essays without looking at students' names since the general impression we have of our students is a potential form of bias. Some teachers mark on the basis of how well they know the student and his or her abilities. It is not uncommon for a teacher to give a higher score to a poorly written script of a good or above-average student by rationalizing that "Juan is really a good student even though he didn't show it on this essay. Maybe he was tired or not feeling well." This is known as the halo effect. Have students put their names on the back of their papers or issue each student a candidate number to prevent this practice.
- 10. Calibrate and recalibrate.**
The best way to achieve inter-rater reliability is to practice. Start early in the academic year by employing the marking criteria and scale in non-test situations. Make students aware from the outset of the criteria and expectations for their work. Reliability can be increased by using multiple marking, which reduces the scope for error that is inherent in a single score.

Extension Activity

Mr. Knott has been asked to contribute a writing prompt for the next midterm test. The course objectives focused on description. Here is the prompt he developed. What aspects of the prompt are good? What aspects of the prompt might prove problematic?

E3 MIDTERM EXAM WRITING SECTION VERSION C	
Student Name: _____	Teacher: _____
I.D. Number: _____	Period/Class Time: _____
WRITING:	DESCRIPTION
	10 pts. total
Write a detailed description of the petrol pump in the picture.	