TESTING

WRITING SKILL

ASSESSING WRITING

Not many centuries ago, writing was a skill that was the exclusive domain of scribes and scholars in educational or religious institutions. Almost every aspect of everyday life for "common" people was carried out orally. Business transactions, records, legal documents, political and military agreements—all were written by specialists whose vocation it was to render language into the written word. Today, the ability to write has become an indispensable skill in our global literate community. Writing skill, at least at rudimentary levels, is a necessary condition for achieving employment in many walks of life and is simply taken for granted in literate cultures.

In the field of second language teaching, only a half-century ago experts were saying that writing was primarily a convention for recording speech and for reinforcing grammatical and lexical features of language. Now we understand the uniqueness of writing as a skill with its own features and conventions. We also fully understand the difficulty of learning to write "well" in any language, even in our own native language. Every educated child in developed countries learns the rudiments of writing in his or her native language, but very few learn to express themselves clearly with logical, well-developed organization that accomplishes an intended purpose. And yet we expect second language learners to write coherent essays with artfully chosen rhetorical and discourse devices!

With such a monumental goal, the job of teaching writing has occupied the attention of papers, articles, dissertations, books, and even separate professional journals exclusively devoted to writing in a second language. I refer specifically to the *Journal of Second Language Writing*; consult the website http://icdweb.cc.purdue.edu/~silvat/jslw/ for information. (For further information on issues and practical techniques in teaching writing, refer to *TBP*, Chapter 19.)

It follows logically that the assessment of writing is no simple task. As you consider assessing students' writing ability, as usual you need to be clear about your objective or criterion. What is it you want to test: handwriting ability? correct spelling? writing sentences that are grammatically correct? paragraph construction? logical development of a main idea? All of these, and more, are possible objectives. And each objective can be assessed through a variety of tasks, which we will examine in this chapter.

Before looking at specific tasks, we must scrutinize the different genres of written language (so that context and purpose are clear), types of writing (so that stages of the development of writing ability are accounted for), and micro- and macroskills of writing (so that objectives can be pinpointed precisely).

GENRES OF WRITTEN LANGUAGE

Chapter 8's discussion of assessment of reading listed more than 50 written language genres. The same classification scheme is reformulated here to include the most common genres that a second language *writer* might produce, within and beyond the requirements of a curriculum. Even though this list is slightly shorter, you should be aware of the surprising multiplicity of options of written genres that second language learners need to acquire.

Genres of writing

1. **Academic writing**
   papers and general subject reports
   essays, compositions
   academically focused journals
   short-answer test responses
   technical reports (e.g., lab reports)
   theses, dissertations

2. **Job-related writing**
   messages (e.g., phone messages)
   letters/emails
   memos (e.g., interoffice)
   reports (e.g., job evaluations, project reports)
   schedules, labels, signs
   advertisements, announcements
   manuals

3. **Personal writing**
   letters, emails, greeting cards, invitations
   messages, notes
   calendar entries, shopping lists, reminders
   financial documents (e.g., checks, tax forms, loan applications)
   forms, questionnaires, medical reports, immigration documents
   diaries, personal journals
   fiction (e.g., short stories, poetry)

TYPES OF WRITING PERFORMANCE

Four categories of written performance that capture the range of written production are considered here. Each category resembles the categories defined for the other three skills, but these categories, as always, reflect the uniqueness of the skill area.

1. Imitative. To produce written language, the learner must attain skills in the fundamental, basic tasks of writing letters, words, punctuation, and very brief sentences. This category includes the ability to spell correctly and to perceive phoneme–grapheme correspondences in the English spelling system. It is a level at which learners are trying to master the mechanics of writing. At this stage, form is the primary if not exclusive focus, while context and meaning are of secondary concern.

2. Intensive (controlled). Beyond the fundamentals of imitative writing are skills in producing appropriate vocabulary within a context, collocations and idioms, and correct grammatical features up to the length of a sentence. Meaning and context are of some importance in determining correctness and appropriateness, but most assessment tasks are more concerned with a focus on form, and are rather strictly controlled by the test design.

3. Responsive. Here, assessment tasks require learners to perform at a limited discourse level, connecting sentences into a paragraph and creating a logically connected sequence of two or three paragraphs. Tasks respond to pedagogical directives, lists of criteria, outlines, and other guidelines. Genres of writing include brief narratives and descriptions, short reports, lab reports, summaries, brief responses to reading, and interpretations of charts or graphs. Under specified conditions, the writer begins to exercise some freedom of choice among alternative forms of expression of ideas. The writer has mastered the fundamentals of sentence-level grammar and is more focused on the discourse conventions that will achieve the objectives of the written text. Form-focused attention is mostly at the discourse level, with a strong emphasis on context and meaning.

4. Extensive. Extensive writing implies successful management of all the processes and strategies of writing for all purposes, up to the length of an essay, a term paper, a major research project report, or even a thesis. Writers focus on achieving a purpose, organizing and developing ideas logically, using details to support or illustrate ideas, demonstrating syntactic and lexical variety, and in many cases, engaging in the process of multiple drafts to achieve a final product. Focus on grammatical form is limited to occasional editing or proofreading of a draft.

MICRO- AND MACROSKILLS OF WRITING

We turn once again to a taxonomy of micro- and macroskills that will assist you in defining the ultimate criterion of an assessment procedure. The earlier microskills apply more appropriately to imitative and intensive types of writing task, while the macroskills are essential for the successful mastery of responsive and extensive writing.

Micro- and macroskills of writing

Microskills

1. Produce graphemes and orthographic patterns of English.
2. Produce writing at an efficient rate of speed to suit the purpose.
3. Produce an acceptable core of words and use appropriate word order patterns.
4. Use acceptable grammatical systems (e.g., tense, agreement, pluralization), patterns, and rules.
5. Express a particular meaning in different grammatical forms.
6. Use cohesive devices in written discourse.

Macroskills

7. Use the rhetorical forms and conventions of written discourse.
8. Appropriately accomplish the communicative functions of written texts according to form and purpose.
9. Convey links and connections between events, and communicate such relations as main idea, supporting idea, new information, given information, generalization, and exemplification.
10. Distinguish between literal and implied meanings when writing.
11. Correctly convey culturally specific references in the context of the written text.
12. Develop and use a battery of writing strategies, such as accurately assessing the audience's interpretation, using prewriting devices, writing with fluency in the first drafts, using paraphrases and synonyms, soliciting peer and instructor feedback, and using feedback for revising and editing.

DESIGNING ASSESSMENT TASKS: IMITATIVE WRITING

With the recent worldwide emphasis on teaching English at young ages, it is tempting to assume that every English learner knows how to handwrite the Roman alphabet. Such is not the case. Many beginning-level English learners, from young children to older adults, need basic training in and assessment of imitative writing: the rudiments of forming letters, words, and simple sentences. We examine this level of writing first.

Tasks in [Hand] Writing Letters, Words, and Punctuation

First, a comment should be made on the increasing use of personal and laptop computers and handheld instruments for creating written symbols. Handwriting has the potential of becoming a lost art as even very young children are more and more likely to use a keyboard to produce writing. Making the shapes of letters and other symbols is now more a question of learning typing skills than of training the muscles

252

of the hands to use a pen or pencil. Nevertheless, for all practical purposes, hand-writing remains a skill of paramount importance within the larger domain of language assessment.

A limited variety of types of tasks are commonly used to assess a person's ability to produce written letters and symbols. A few of the more common types are described here.

1. Copying. There is nothing innovative or modern about directing a test-taker to copy letters or words. The test-taker will see something like the following:

Handwriting letters, words, and punctuation marks

> *The test-taker reads:* Copy the following words in the spaces given:
>
> bit   bet   bat   but   Oh?   Oh!
>
> ___   ___   ___   ___   ___   ___
>
> bin   din   gin   pin   Hello, John.
>
> ___   ___   ___   ___   ___

2. Listening cloze selection tasks. These tasks combine dictation with a written script that has a relatively frequent deletion ratio (every fourth or fifth word, perhaps). The test sheet provides a list of missing words from which the test-taker must select. The purpose at this stage is not to test spelling but to give practice in writing. To increase the difficulty, the list of words can be deleted, but then spelling might become an obstacle. Probes look like this:

Listening cloze selection task

> *Test-takers hear:*
> Write the missing word in each blank. Below the story is a list of words to choose from.
>
> Have you ever visited San Francisco? It is a very nice city. It is cool in the summer and warm in the winter. I like the cable cars and bridges.
>
> *Test-takers see:*
>
> Have _____ ever visited San Francisco? It _____ a very nice _____. It is _____ in _____ summer and _____ in the winter. I _____ the cable cars _____ bridges.
>
> is   you   cool   city
> like   and   warm   the

3. *Picture-cued tasks.* Familiar pictures are displayed, and test-takers are told to write the word that the picture represents. Assuming no ambiguity in identifying the picture (cat, hat, chair, table, etc.), no reliance is made on aural comprehension for successful completion of the task.

4. *Form completion tasks.* A variation on pictures is the use of a simple form (registration, application, etc.) that asks for name, address, phone number, and other data. Assuming, of course, that prior classroom instruction has focused on filling out such forms, this task becomes an appropriate assessment of simple tasks such as writing one's name and address.

5. *Converting numbers and abbreviations to words.* Some tests have a section on which numbers are written—for example, hours of the day, dates, or schedules—and test-takers are directed to write out the numbers. This task can serve as a reasonably reliable method to stimulate handwritten English. It lacks authenticity, however, in that people rarely write out such numbers (except in writing checks), and it is more of a reading task (recognizing numbers) than a writing task. If you plan to use such a method, be sure to specify exactly what the criterion is, and then proceed with some caution. Converting abbreviations to words is more authentic: we actually do have occasions to write out days of the week, months, and words like *street, boulevard, telephone,* and *April* (months of course are often abbreviated with numbers). Test tasks may take this form:

Writing numbers and abbreviations

Test-takers hear: Fill in the blanks with words.

Test-takers see:

9:00	_____	5:45	_____
Tues.	_____	5/3	_____
726 S. Main St.	_____		

Spelling Tasks and Detecting Phoneme–Grapheme Correspondences

A number of task types are in popular use to assess the ability to spell words correctly and to process phoneme–grapheme correspondences.

1. *Spelling tests.* In a traditional, old-fashioned spelling test, the teacher dictates a simple list of words, one word at a time, followed by the word in a sentence, repeated again, with a pause for test-takers to write the word. Scoring emphasizes correct spelling. You can help to control for listening errors by choosing words that

the students have encountered before—words that they have spoken or heard in their class.

2. *Picture-cued tasks.* Pictures are displayed with the objective of focusing on familiar words whose spelling may be unpredictable. Items are chosen according to the objectives of the assessment, but this format is an opportunity to present some challenging words and word pairs: *boot/book, read/reed, bit/bite,* etc.

3. *Multiple-choice techniques.* Presenting words and phrases in the form of a multiple-choice task risks crossing over into the domain of assessing reading, but if the items have a follow-up writing component, they can serve as formative reinforcement of spelling conventions. They might be more challenging with the addition of homonyms (see item #3 below). Here are some examples.

Multiple-choice reading-writing spelling tasks

Test-takers read:

Choose the word with the correct spelling to fit the sentence, then write the word in the space provided.

1. He washed his hands with _____.
   A. soap
   B. sope
   C. sop
   D. soup

2. I tried to stop the car, but the _____ didn't work.
   A. braicks
   B. brecks
   C. brakes
   D. bracks

3. The doorbell rang, but when I went to the door, no one was _____.
   A. their
   B. there
   C. they're
   D. thair

---

4. *Matching phonetic symbols.* If students have become familiar with the phonetic alphabet, they could be shown phonetic symbols and asked to write the correctly spelled word alphabetically. This works best with letters that do not have one-to-one correspondence with the phonetic symbol (e.g., /æ/ and a). In the sample below, the answers, which of course do not appear on the test sheet, are included in brackets for your reference.

Converting phonetic symbols

Test-takers read:

In each of the following words, a letter or combination of letters has been written in a phonetic symbol. Write the word using the regular alphabet.

1. tea /tʃ/ er   _____   [teacher]
2. d /e/   _____   [day]
3. /ð/ is   _____   [this]
4. n /au/   _____   [now]
5. l /aɪ/ /k/   _____   [like]
6. c /æ/ t   _____   [cat]

Such a task risks confusing students who don't recognize the phonetic alphabet or use it in their daily routine. Opinion is mixed on the value of using phonetic symbols at the literacy level. Some claim it helps students to perceive the relationship between phonemes and graphemes. Others caution against using yet another system of symbols when the alphabet already poses a challenge, especially for adults for whom English is the only language they have learned to read or write.

DESIGNING ASSESSMENT TASKS: INTENSIVE (CONTROLLED) WRITING

This next level of writing is what second language teacher training manuals have for decades called **controlled** writing. It may also be thought of as form-focused writing, grammar writing, or simply guided writing. A good deal of writing at this level is **display writing** as opposed to **real writing**: students produce language to display their competence in grammar, vocabulary, or sentence formation, and not necessarily to convey meaning for an authentic purpose. The traditional grammar/vocabulary test has plenty of display writing in it, since the response mode demonstrates only the test-taker's ability to combine or use words correctly. No new information is passed on from one person to the other.

Dictation and Dicto-Comp

In Chapter 6, dictation was described as an assessment of the integration of listening and writing, but it was clear that the primary skill being assessed is listening. Because of its response mode, however, it deserves a second mention in this chapter. Dictation is simply the rendition in writing of what one hears aurally, so it could be classified as an *imitative* type of writing, especially since a proportion of the test-taker's performance centers on correct spelling. Also, because the test-taker must listen to stretches of discourse and in the process insert punctuation, dictation of a

paragraph or more can arguably be classified as a controlled or intensive form of writing. (For a further explanation on administering a dictation, consult Chapter 6, pages 131–132.)

A form of controlled writing related to dictation is a **dicto-comp**. Here, a paragraph is read at normal speed, usually two or three times; then the teacher asks students to rewrite the paragraph from the best of their recollection. In one of several variations of the dicto-comp technique, the teacher, after reading the passage, distributes a handout with key words from the paragraph, in sequence, as cues for the students. In either case, the dicto-comp is genuinely classified as an intensive, if not a responsive, writing task. Test-takers must internalize the content of the passage, remember a few phrases and lexical items as key words, then recreate the story in their own words.

Grammatical Transformation Tasks

In the heyday of structural paradigms of language teaching with slot-filler techniques and slot substitution drills, the practice of making grammatical transformations—orally or in writing—was very popular. To this day, language teachers have also used this technique as an assessment task, ostensibly to measure grammatical competence. Numerous versions of the task are possible:

- Change the tenses in a paragraph.
- Change full forms of verbs to reduced forms (contractions).
- Change statements to *yes/no* or *wh*-questions.
- Change questions into statements.
- Combine two sentences into one using a relative pronoun.
- Change direct speech to indirect speech.
- Change from active to passive voice.

The list of possibilities is almost endless. The tasks are virtually devoid of any meaningful value. Sometimes test designers attempt to add authenticity by providing a context ("Today Doug is doing all these things. Tomorrow he will do the same things again. Write about what Doug will do tomorrow by using the future tense."), but this is just a backdrop for a written substitution task. On the positive side, grammatical transformation tasks are easy to administer and are therefore practical, quite high in scorer reliability, and arguably tap into a knowledge of grammatical *forms* that will be performed through writing. If you are only interested in a person's ability to produce the forms, then such tasks may prove to be justifiable.

Picture-Cued Tasks

A variety of picture-cued controlled tasks have been used in English classrooms around the world. The main advantage in this technique is in detaching the almost ubiquitous reading and writing connection and offering instead a nonverbal means to stimulate written responses.

257

1. Short sentences. A drawing of some simple action is shown; the test-taker writes a brief sentence.

Picture-cued sentence writing (Brown, 1999, p. 40)
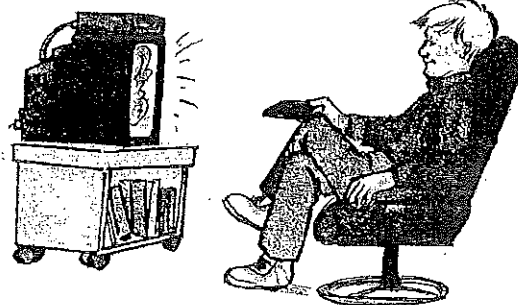
Test-takers see the following pictures:

1.

2.

3.

Test-takers read: What is the woman doing?
What is the man doing?
What is the boy doing?

Test-takers write:

1. *She is eating. She is eating her dinner. She is holding a spoon,* etc.

2. Picture description. A somewhat more complex picture may be presented showing, say, a person reading on a couch, a cat under a table, books and pencils on the table, chairs around the table, a lamp next to the couch, and a picture on the wall over the couch (see Chapter 8, page 192). Test-takers are asked to describe the picture using four of the following prepositions: *on, over, under, next to, around.* As long as the prepositions are used appropriately, the criterion is considered to be met.

3. Picture sequence description. A sequence of three to six pictures depicting a story line can provide a suitable stimulus for written production. The pictures must be simple and unambiguous because an open-ended task at the selective level would give test-takers too many options. If writing the correct grammatical form of a verb is the only criterion, then some test items might include the simple form of the verb

below the picture. The time sequence in the following task is intended to give writers some cues.

Picture-cued story sequence (Brown, 1999, p. 43)

Test-takers see:
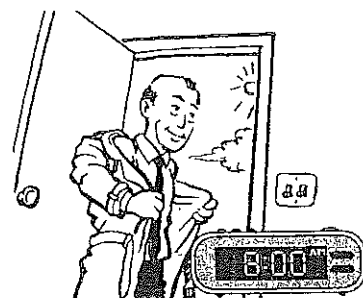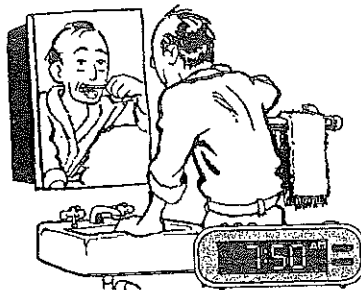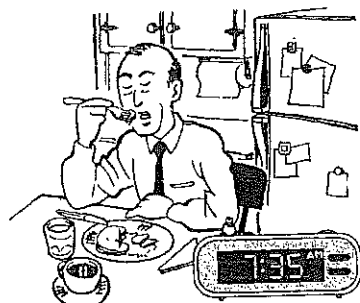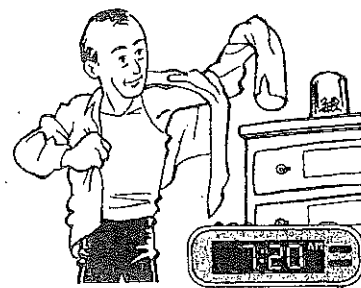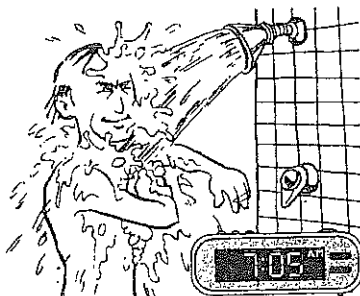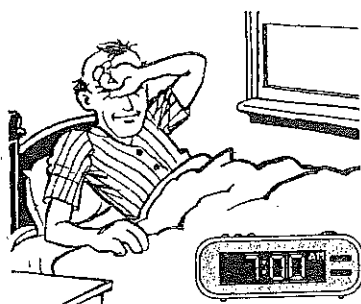
Test-takers read: Describe the man's morning routine in six sentences.

Test-takers write:
He gets up at seven o'clock.
He takes a shower at 7:05.
At 7:20, he gets dressed.
Then he eats breakfast.
About 7:50 he brushes his teeth.
He leaves the house at eight.

While these kinds of tasks are designed to be controlled, even at this very simple level, a few different correct responses can be made for each item in the sequence. If your criteria in this task are both lexical and grammatical choice, then you need to design a rating scale to account for variations between completely right and completely wrong in both categories.

Scoring scale for controlled writing

2	Grammatically and lexically correct.
1	Either grammar or vocabulary is incorrect, but not both.
0	Both grammar and vocabulary are incorrect.

The following are some test-takers' responses to the first picture:

> He gets up at 7.
> He get up at 7.
> He is getting up at 7.
> He wakes seven o'clock.
> The man is arise at seven.
> He sleeps at seven o'clock.
> Sleeps on morning.

How would you rate each response? With the scoring scale above, the first response is a "2," the next five responses are a "1," and the last earns a zero.

Vocabulary Assessment Tasks

Most vocabulary study is carried out through reading. A number of assessments of reading recognition of vocabulary were discussed in the previous chapter: multiple-choice techniques, matching, picture-cued identification, cloze techniques, guessing the meaning of a word in context, etc. The major techniques used to assess vocabulary are (a) defining and (b) using a word in a sentence. The latter is the more authentic, but even that task is constrained by a contrived situation in which the test-taker, usually in a matter of seconds, has to come up with an appropriate sentence, which may or may not indicate that the test-taker "knows" the word.

Read (2000) suggested several types of items for assessment of basic knowledge of the meaning of a word, collocational possibilities, and derived morphological forms. His example centered on the word *interpret*, as follows:

Vocabulary writing tasks (Read, 2000, p. 179)

Test-takers read:

1. Write two sentences, A and B. In each sentence, use the two words given.

   A. *interpret, experiment* _____.

   B. *interpret, language* _____.

2. Write three words that can fit in the blank.

   To interpret a(n) _____   i. _____

                                     ii. _____

                                     iii. _____

3. Write the correct ending for the word in each of the following sentences:

   Someone who interprets is an interpret _____.

   Something that can be interpreted is interpret _____.

   Someone who interprets gives an interpret _____.

Vocabulary assessment is clearly form-focused in the above tasks, but the procedures are creatively linked by means of the target word, its collocations, and its morphological variants. At the responsive and extensive levels, where learners are called upon to create coherent paragraphs, performance obviously becomes more authentic, and lexical choice is one of several possible components of the evaluation of extensive writing.

Ordering Tasks

One task at the sentence level may appeal to those who are fond of word games and puzzles: ordering (or reordering) a scrambled set of words into a correct sentence. Here is the way the item format appears.

Reordering words in a sentence

Test-takers read:

Put the words below into the correct order to make a sentence:

1. cold / winter / is / weather / the / in / the

2. studying / what / you / are

3. next / clock / the / the / is / picture / to

Test-takers write:

1. *The weather is cold in the winter.*

2. *What are you studying?*

3. *The clock is next to the picture.*

While this somewhat inauthentic task generates writing performance and may be said to tap into grammatical word-ordering rules, it presents a challenge to test-takers whose learning styles do not dispose them to logical-mathematical problem solving. If sentences are kept very simple (such as #2) with perhaps no more than four or five words, if only one possible sentence can emerge, and if students have practiced the technique in class, then some justification emerges. But once again, as in so many writing techniques, this task involves as much, if not more, reading performance as writing.

Short-Answer and Sentence Completion Tasks

Some types of short-answer tasks were discussed in Chapter 8 because of the heavy participation of reading performance in their completion. Such items range from very simple and predictable to somewhat more elaborate responses. Look at the range of possibilities.

Limited response writing tasks

Test-takers see:

1. Alicia: Who's that?
   Tony: _____ Gina.
   Alicia: Where's she from?
   Tony: _____ Italy.

2. Jennifer: _____ ?
   Kathy: I'm studying English.

3. Restate the following sentences in your own words, using the underlined word. You may need to change the meaning of the sentence a little.
   3a. I never miss a day of school.   <u>always</u>
   3b. I'm pretty healthy most of the time.   <u>seldom</u>
   3c. I play tennis twice a week.   <u>sometimes</u>

4. You are in the kitchen helping your roommate cook. You need to ask questions about quantities. Ask a question using *how much* (#4a) and a question using *how many* (#4b), using nouns like *sugar, pounds, flour, onions, eggs, cups.*
   4a. _____ .
   4b. _____ .

5. Look at the schedule of Roberto's week. Write two sentences describing what Roberto does, using the words *before* (#5a) and *after* (#5b).
   5a. _____ .
   5b. _____ .

6. Write three sentences describing your preferences: #6a: a big, expensive car or a small, cheap car; #6b: a house in the country or an apartment in the city; #6c: money or good health.
   6a. _____ .
   6b. _____ .
   6c. _____ .

The reading-writing connection is apparent in the first three item types but has less of an effect in the last three, where reading is necessary in order to understand the directions but is not crucial in creating sentences. Scoring on a 2-1-0 scale (as described above) may be the most appropriate way to avoid self-arguing about the appropriateness of a response.

ISSUES IN ASSESSING RESPONSIVE AND EXTENSIVE WRITING

Responsive writing creates the opportunity for test-takers to offer an array of possible creative responses within a pedagogical or assessment framework: test-takers are "responding" to a prompt or assignment. Freed from the strict control of intensive

writing, learners can exercise a number of options in choosing vocabulary, grammar, and discourse, but with some constraints and conditions. Criteria now begin to include the discourse and rhetorical conventions of paragraph structure and of connecting two or three such paragraphs in texts of limited length. The learner is responsible for accomplishing a purpose in writing, for developing a sequence of connected ideas, and for empathizing with an audience.

The genres of text that are typically addressed here are

- short reports (with structured formats and conventions);
- responses to the reading of an article or story;
- summaries of articles or stories;
- brief narratives or descriptions; and
- interpretations of graphs, tables, and charts.

It is here that writers become involved in the art (and science) of *composing*, or real writing, as opposed to display writing.

Extensive, or "free," writing, which is amalgamated into our discussion here, takes all the principles and guidelines of responsive writing and puts them into practice in longer texts such as full-length essays, term papers, project reports, and theses and dissertations. In extensive writing, however, the writer has been given even more freedom to choose: topics, length, style, and perhaps even conventions of formatting are less constrained than in the typical responsive writing exercise. At this stage, all the rules of effective writing come into play, and the second language writer is expected to meet all the standards applied to native language writers.

Both responsive and extensive writing tasks are the subject of some classic, widely debated assessment issues that take on a distinctly different flavor from those at the lower-end production of writing.

1. Authenticity. Authenticity is a trait that is given special attention: if test-takers are being asked to perform a task, its face and content validity need to be assured in order to bring out the best in the writer. A good deal of writing performance in academic contexts is constrained by the pedagogical necessities of establishing the basic building blocks of writing; we have looked at assessment techniques that address those foundations. But once those fundamentals are in place, the would-be writer is ready to fly out of the protective nest of the writing classroom and assume his or her own voice. Offering that freedom to learners requires the setting of authentic real-world contexts in which to write. The teacher becomes less of an instructor and more of a coach or facilitator. Assessment therefore is typically formative, not summative, and positive washback is more important than practicality and reliability.

2. Scoring. Scoring is the thorniest issue at these final two stages of writing. With so many options available to a learner, each evaluation by a test administrator needs to be finely attuned not just to how the writer strings words together (the *form*) but also to what the writer is saying (the *function* of the text). The quality of writing (its impact and effectiveness) becomes as important, if not more important,

than all the nuts and bolts that hold it together. How are you to score such creative production, some of which is more artistic than scientific? A discussion of different scoring options will continue below, followed by a reminder that responding and editing are nonscoring options that yield washback to the writer.

3. Time. Yet another assessment issue surrounds the unique nature of writing: it is the only skill in which the language producer is not necessarily constrained by *time,* which implies the freedom to process multiple drafts before the text becomes a finished product. Like a sculptor creating an image, the writer can take an initial rough conception of a text and continue to refine it until it is deemed presentable to the public eye. Virtually all *real* writing of prose texts presupposes an extended time period for it to reach its final form, and therefore the revising and editing processes are implied. Responsive writing, along with the next category of extensive writing, often relies on this essential drafting process for its ultimate success.

How do you assess writing ability within the confines of traditional, formal assessment procedures that are almost always, by logistical necessity, *timed?* We have a whole testing industry that has based large-scale assessment of writing on the premise that the **timed impromptu** format is a valid method of assessing writing ability. Is this an authentic format? Can a language learner—or a native speaker, for that matter—adequately perform writing tasks within the confines of a brief timed period of composition? Is that hastily written product an appropriate reflection of what that same test-taker might produce after several drafts of the same work? Does this format favor fast writers at the expense of slower but possibly equally good or better writers? Alderson (2002) and Weigle (2002) both cited this as one of the most pressing unresolved issues in the assessment of writing today. We will return to this question below.

Because of the complexity of assessing responsive and extensive writing, the discussion that ensues will now have a different look from the one used in the previous three chapters. Four major topics will be addressed: (1) a few fundamental task types at the lower (responsive) end of the continuum of writing at this level; (2) a description and analysis of the *Test of Written English*® as a typical timed impromptu test of writing; (3) a survey of methods of scoring and evaluating writing production; and (4) a discussion of the assessment qualities of editing and responding to a series of writing drafts.

DESIGNING ASSESSMENT TASKS: RESPONSIVE AND EXTENSIVE WRITING

In this section we consider both responsive and extensive writing tasks. They will be regarded here as a continuum of possibilities ranging from lower-end tasks whose complexity exceeds those in the previous category of intensive or controlled writing, through more open-ended tasks such as writing short reports, essays, summaries, and responses, up to texts of several pages or more.

Paraphrasing

One of the more difficult concepts for second language learners to grasp is paraphrasing. The initial step in teaching paraphrasing is to ensure that learners understand the importance of paraphrasing: to say something in one's own words, to avoid plagiarizing, to offer some variety in expression. With those possible motivations and purposes in mind, the test designer needs to elicit a paraphrase of a sentence or paragraph, usually not more.

Scoring of the test-taker's response is a judgment call in which the criterion of conveying the same or similar message is primary, with secondary evaluations of discourse, grammar, and vocabulary. Other components of analytic or holistic scales (see discussion below, page 242) might be considered as criteria for an evaluation. Paraphrasing is more often a part of informal and formative assessment than of formal, summative assessment, and therefore student responses should be viewed as opportunities for teachers and students to gain positive washback on the art of paraphrasing.

Guided Question and Answer

Another lower-order task in this type of writing, which has the pedagogical benefit of guiding a learner without dictating the form of the output, is a guided question-and-answer format in which the test administrator poses a series of questions that essentially serve as an outline of the emergent written text. In the writing of a narrative that the teacher has already covered in a class discussion, the following kinds of questions might be posed to stimulate a sequence of sentences:

Guided writing stimuli

---

1. Where did this story take place? [setting]
2. Who were the people in the story? [characters]
3. What happened first? and then? and then? [sequence of events]
4. Why did _____ do _____? [reasons, causes]
5. What did _____ think about _____? [opinion]
6. What happened at the end? [climax]
7. What is the moral of this story? [evaluation]

Guided writing texts, which may be as long as two or three paragraphs, may be scored on either an analytic or a holistic scale (discussed below). Guided writing prompts like these are less likely to appear on a formal test and more likely to serve as a way to prompt initial drafts of writing. This first draft can then undergo the editing and revising stages discussed in the next section of this chapter.

A variation on using guided questions is to prompt the test-taker to write from an outline. The outline may be self-created from earlier reading and/or discussion, or, which is less desirable, be provided by the teacher or test administrator. The outline helps to guide the learner through a presumably logical development of ideas that have been given some forethought. Assessment of the resulting text follows the same criteria listed below (#3 in the next section, paragraph construction tasks).

Paragraph Construction Tasks

The participation of reading performance is inevitable in writing effective paragraphs. To a great extent, writing is the art of emulating what one reads. You read an effective paragraph; you analyze the ingredients of its success; you emulate it. Assessment of paragraph development takes on a number of different forms:

1. Topic sentence writing. There is no cardinal rule that says every paragraph must have a topic sentence, but the stating of a topic through the lead sentence (or a subsequent one) has remained as a tried-and-true technique for teaching the concept of a paragraph. Assessment thereof consists of

- specifying the writing of a topic sentence,
- scoring points for its presence or absence, and
- scoring and/or commenting on its effectiveness in stating the topic.

2. Topic development within a paragraph. Because paragraphs are intended to provide a reader with "clusters" of meaningful, connected thoughts or ideas, another stage of assessment is development of an idea within a paragraph. Four criteria are commonly applied to assess the quality of a paragraph:

- the clarity of expression of ideas
- the logic of the sequence and connections
- the cohesiveness or unity of the paragraph
- the overall effectiveness or impact of the paragraph as a whole

3. Development of main and supporting ideas across paragraphs. As writers string two or more paragraphs together in a longer text (and as we move up the continuum from responsive to extensive writing), the writer attempts to articulate a thesis or main idea with clearly stated supporting ideas. These elements can be considered in evaluating a multi-paragraph essay:

- addressing the topic, main idea, or principal purpose
- organizing and developing supporting ideas
- using appropriate details to undergird supporting ideas
- showing facility and fluency in the use of language
- demonstrating syntactic variety

Strategic Options

Developing main and supporting ideas is the goal for the writer attempting to create an effective text, whether a short one- to two-paragraph one or an extensive one of several pages. A number of strategies are commonly taught to second language writers to accomplish their purposes. Aside from strategies of freewriting, outlining, drafting, and revising, writers need to be aware of the task that has been demanded and to focus on the genre of writing and the expectations of that genre.

1. Attending to task. In responsive writing, the context is seldom completely open-ended: a task has been defined by the teacher or test administrator, and the writer must fulfill the criterion of the task. Even in extensive writing of longer texts, a set of directives has been stated by the teacher or is implied by the conventions of the genre. Four types of tasks are commonly addressed in academic writing courses: compare/contrast, problem/solution, pros/cons, and cause/effect. Depending on the genre of the text, one or more of these task types will be needed to achieve the writer's purpose. If students are asked, for example, to "agree or disagree with the author's statement," a likely strategy would be to cite pros and cons and then take a stand. A task that asks students to argue for one among several political candidates in an election might be an ideal compare-and-contrast context, with an appeal to problems present in the constituency and the relative value of candidates' solutions. Assessment of the fulfillment of such tasks could be formative and informal (comments in marginal notes, feedback in a conference in an editing/revising stage), but the product might also be assigned a holistic or analytic score.

2. Attending to genre. The genres of writing that were listed at the beginning of this chapter provide some sense of the many varieties of text that may be produced by a second language learner in a writing curriculum. Another way of looking at the strategic options open to a writer is the extent to which both the constraints and the opportunities of the genre are exploited. Assessment of any writing necessitates attention to the conventions of the genre in question. Assessment of the more common genres may include the following criteria, along with chosen factors from the list in item #3 (main and supporting ideas) above:

Reports (Lab Reports, Project Summaries, Article/Book Reports, etc.)
- conform to a conventional format (for this case, field)
- convey the purpose, goal, or main idea
- organize details logically and sequentially
- state conclusions or findings
- use appropriate vocabulary and jargon for the specific case

Summaries of Readings/Lectures/Videos
- effectively capture the main and supporting ideas of the original
- maintain objectivity in reporting
- use writer's own words for the most part

267

- use quotations effectively when appropriate
- omit irrelevant or marginal details
- conform to an expected length

Responses to Readings/Lectures/Videos
- accurately reflect the message or meaning of the original
- appropriately select supporting ideas to respond to
- express the writer's own opinion
- defend or support that opinion effectively
- conform to an expected length

Narration, Description, Persuasion/Argument, and Exposition
- follow expected conventions for each type of writing
- convey purpose, goal, or main idea
- use effective writing strategies
- demonstrate syntactic variety and rhetorical fluency

Interpreting Statistical, Graphic, or Tabular Data
- provides an effective global, overall description of the data
- organizes the details in clear, logical language
- accurately conveys details
- appropriately articulates relationships among elements of the data
- conveys specialized or complex data comprehensibly to a lay reader
- interprets beyond the data when appropriate

Library Research Paper
- states purpose or goal of the research
- includes appropriate citations and references in correct format
- accurately represents others' research findings
- injects writer's own interpretation, when appropriate, and justifies it
- includes suggestions for further research
- sums up findings in a conclusion

TEST OF WRITTEN ENGLISH (TWE®)

One of a number of internationally available standardized tests of writing ability is the *Test of Written English (TWE)*. Established in 1986, the TWE has gained a reputation as a well-respected measure of written English, and a number of research articles support its validity (Frase et al., 1999; Hale et al., 1996; Longford, 1996; Myford et al., 1996). In 1998, a computer-delivered version of the TWE was incorporated into the standard computer-based TOEFL and simply labeled as the "writing" section of the TOEFL. The TWE is still offered as a separate test especially where only the paper-based TOEFL is available. Correlations between the TWE and TOEFL scores (before TWE became a standard part of TOEFL) were consistently high, ranging from .57 to

.69 over 10 test administrations from 1993 to 1995. Data on the TWE are provided at the end of this section.

The TWE is in the category of a **timed impromptu** test in that test-takers are under a 30-minute time limit and are not able to prepare ahead of time for the topic that will appear. Topics are prepared by a panel of experts following specifications for topics that represent commonly used discourse and thought patterns at the university level. Here are some sample topics published on the TWE website.

Sample TWE® topics

---

1. Some people say that the best preparation for life is learning to work with others and be cooperative. Others take the opposite view and say that learning to be competitive is the best preparation. Discuss these positions, using concrete examples of both. Tell which one you agree with and explain why.
2. Some people believe that automobiles are useful and necessary. Others believe that automobiles cause problems that affect our health and well-being. Which position do you support? Give specific reasons for your answer.
3. Do you agree or disagree with the following statement?

   Teachers should make learning enjoyable and fun for their students.

   Use reasons and specific examples to support your opinion.

Test preparation manuals such as Deborah Phillips's *Longman Introductory Course for the TOEFL Test* (2001) advise TWE test-takers to follow six steps to maximize success on the test:

1. Carefully identify the topic.
2. Plan your supporting ideas.
3. In the introductory paragraph, restate the topic and state the organizational plan of the essay.
4. Write effective supporting paragraphs (show transitions, include a topic sentence, specify details).
5. Restate your position and summarize in the concluding paragraph.
6. Edit sentence structure and rhetorical expression.

The scoring guide for the TWE (see Table 9.1) follows a widely accepted set of specifications for a holistic evaluation of an essay (see below for more discussion of holistic scoring). Each point on the scoring system is defined by a set of statements that address topic, organization and development, supporting ideas, facility (fluency, naturalness, appropriateness) in writing, and grammatical and lexical correctness and choice.

Table 9.1. Test of Written English Scoring Guide

6   **Demonstrates clear competence in writing on both the rhetorical and syntactic levels, though it may have occasional errors.**

A paper in this category
- effectively addresses the writing task.
- is well organized and well developed.
- uses clearly appropriate details to support a thesis or illustrate ideas.
- displays consistent facility in the use of language.
- demonstrates syntactic variety and appropriate word choice.

5   **Demonstrates competence in writing on both the rhetorical and syntactic levels, though it will probably have occasional errors.**

A paper in this category
- may address some parts of the task more effectively than others.
- is generally well organized and developed.
- uses details to support a thesis or illustrate an idea.
- displays facility in the use of language.
- demonstrates some syntactic variety and range of vocabulary.

4   **Demonstrates minimal competence in writing on both the rhetorical and syntactic levels.**

A paper in this category
- addresses the writing topic adequately but may slight parts of the task.
- is adequately organized and developed.
- uses some details to support a thesis or illustrate an idea.
- demonstrates adequate but possibly inconsistent facility with syntax and usage.
- may contain some errors that occasionally obscure meaning.

3   **Demonstrates some developing competence in writing, but it remains flawed on either the rhetorical or syntactic level, or both.**

A paper in this category may reveal one or more of the following weaknesses:
- inadequate organization or development
- inappropriate or insufficient details to support or illustrate generalizations
- a noticeably inappropriate choice of words or word forms
- an accumulation of errors in sentence structure and/or usage.

2   **Suggests incompetence in writing.**

A paper in this category is seriously flawed by one or more of the following weaknesses:
- serious disorganization or underdevelopment
- little or no detail, or irrelevant specifics
- serious and frequent errors in sentence structure or usage
- serious problems with focus.

1   **Demonstrates incompetence in writing.**

A paper in this category
- may be incoherent.
- may be undeveloped.
- may contain severe and persistent writing errors.

0   A paper is rated 0 if it contains no response, merely copies the topic, is off-topic, is written in a foreign language, or consists only of keystroke characters.

Each essay is scored by two trained readers working independently. The final score assigned is the mean of the two independent ratings. The test-taker can achieve a score ranging from 1 to 6 with possible half-points (e.g., 4.5, 5.5) in between. In the case of a discrepancy of more than one point, a third reader resolves the difference. Discrepancy rates are extremely low, usually ranging from 1 to 2 percent per reading.

It is important to put tests like the TWE in perspective. Timed impromptu tests have obvious limitations if you are looking for an authentic sample of performance in a real-world context. How many times in real-world situations (other than in academic writing classes!) will you be asked to write an essay in 30 minutes? Probably never, but the TWE and other standardized timed tests are not intended to mirror the real world. Instead, they are intended to elicit a sample of writing performance that will be *indicative* of a person's writing ability in the real world. TWE designers sought to validate a feasible timed task that would be manageable within their constraints and at the same time offer useful information about the test-taker.

How does the Educational Testing Service justify the TWE as such an indicator? Research by Hale et al. (1996) showed that the prompts used in the TWE approximate writing tasks assigned in 162 graduate and undergraduate courses across several disciplines in eight universities. Another study (Golub-Smith et al., 1993) ascertained the reliabilities across several types of prompts (e.g., compare/contrast vs. chart-graph interpretation). Both Myford et al. (1996) and Longford (1996) studied the reliabilities of judges' ratings. The question of whether a mere 30-minute time period is sufficient to elicit a sufficient sample of a test-taker's writing was addressed by Hale (1992). Henning and Cascallar (1992) conducted a large-scale study to assess the extent to which TWE performance taps into the communicative competence of the test-taker. The upshot of this research—which is updated regularly—is that the TWE (which adheres to a high standard of excellence in standardized testing) is, within acceptable standard error ranges, a remarkably accurate indicator of writing ability.

The flip side of this controversial coin reminds us that standardized tests are indicators, not fail-safe, infallible measures of competence. Even though we might need TWE scores for the *administrative* purposes of admissions or placement, we should not rely on such tests for *instructional* purposes (see Cohen, 1994). No one would suggest that such 30-minute writing tests offer constructive feedback to the student, nor do they provide the kind of formative assessment that a process approach to writing brings. Tests like the TWE are administrative necessities in a world where hundreds or thousands of applicants must be evaluated by some means short of calculating their performance across years of instruction in academic writing.

The convenience of the TWE should not lull administrators into believing that TWEs and TOEFLs and the like are the only measures that should be applied to students. It behooves admissions and placement officers worldwide to offer secondary measures of writing ability to those test-takers who

a. are on the threshold of a minimum score,
b. may be disabled by highly time-constrained or anxiety-producing situations,
c. could be culturally disadvantaged by a topic or situation, and/or
d. (in the case of computer-based writing) have had few opportunities to compose on a computer.

While timed impromptu tests suffer from a lack of authenticity and put test-takers into an artificially time-constrained context, they nevertheless offer interesting, relevant information for an important but narrow range of administrative purposes. The classroom offers a much wider set of options for creating real-world writing purposes and contexts. The classroom becomes the locus of extended hard work and effort for building the skills necessary to create written production. The classroom provides a setting for writers, in a process of multiple drafts and revisions, to create a final, publicly acceptable product. And the classroom is a place where learners can take all the small steps, at their own pace, toward becoming proficient writers. For your reference, following is some information on the TWE:

Test of Written English (TWE®)

Producer:	Educational Testing Service (ETS), Princeton, NJ
Objective:	To test written expression
Primary market:	Almost exclusively U.S. universities and colleges for admission purposes
Type:	Computer-based, with the TOEFL. A traditional paper-based (PB) version is also available separately.
Response modes:	Written essay
Specifications:	(see above, in this section)
Time allocation:	30 minutes
Internet access:	http://www.toefl.org/educator/edabttwe.html

SCORING METHODS FOR RESPONSIVE AND EXTENSIVE WRITING

At responsive and extensive levels of writing, three major approaches to scoring writing performance are commonly used by test designers: holistic, primary trait, and analytical. In the first method, a single score is assigned to an essay, which represents a reader's general overall assessment. Primary trait scoring is a variation of the holistic method in that the achievement of the primary purpose, or trait, of an essay is the only factor rated. Analytical scoring breaks a test-taker's written text down into a number of subcategories (organization, grammar, etc.) and gives a separate rating for each.

Holistic Scoring

The TWE scoring scale above is a prime example of **holistic** scoring. In Chapter 7, a rubric for scoring oral production holistically was presented. Each point on a holistic scale is given a systematic set of descriptors, and the reader-evaluator matches an overall impression with the descriptors to arrive at a score. Descriptors usually (but not always) follow a prescribed pattern. For example, the first descriptor across all score categories may address the quality of task achievement, the second may deal with organization, the third with grammatical or rhetorical considerations, and so on. Scoring, however, is truly holistic in that those subsets are not quantitatively added up to yield a score.

Advantages of holistic scoring include

- fast evaluation,
- relatively high inter-rater reliability,
- the fact that scores represent "standards" that are easily interpreted by lay persons,
- the fact that scores tend to emphasize the writer's strengths (Cohen, 1994, p. 315), and
- applicability to writing across many different disciplines.

Its disadvantages must also be weighed into a decision on whether to use holistic scoring:

- One score masks differences across the subskills within each score.
- No diagnostic information is available (no washback potential).
- The scale may not apply equally well to all genres of writing.
- Raters need to be extensively trained to use the scale accurately.

In general, teachers and test designers lean toward holistic scoring only when it is expedient for administrative purposes. As long as trained evaluators are in place, differentiation across six levels may be quite adequate for admission into an institution or placement into courses. For classroom instructional purposes, holistic scores provide very little information. In most classroom settings where a teacher wishes to adapt a curriculum to the needs of a particular group of students, much more differentiated information across subskills is desirable than is provided by holistic scoring.

Primary Trait Scoring

A second method of scoring, *primary trait*, focuses on "how well students can write within a narrowly defined range of discourse" (Weigle, 2002, p. 110). This type of scoring emphasizes the task at hand and assigns a score based on the effectiveness of the text's achieving that one goal. For example, if the purpose or function of

an essay is to *persuade* the reader to do something, the score for the writing would rise or fall on the accomplishment of that function. If a learner is asked to exploit the imaginative function of language by expressing personal feelings, then the response would be evaluated on that feature alone.

For rating the primary trait of the text, Lloyd-Jones (1977) suggested a four-point scale ranging from zero (no response or fragmented response) to 4 (the purpose is unequivocally accomplished in a convincing fashion). It almost goes without saying that organization, supporting details, fluency, syntactic variety, and other features will implicitly be evaluated in the process of offering a primary trait score. But the advantage of this method is that it allows both writer and evaluator to focus on function. In summary, a primary trait score would assess

- the accuracy of the account of the original (summary),
- the clarity of the steps of the procedure and the final result (lab report),
- the description of the main features of the graph (graph description), and
- the expression of the writer's opinion (response to an article).

Analytic Scoring

For classroom instruction, holistic scoring provides little washback into the writer's further stages of learning. Primary trait scoring focuses on the principal function of the text and therefore offers some feedback potential, but no washback for any of the aspects of the written production that enhance the ultimate accomplishment of the purpose. Classroom evaluation of learning is best served through **analytic scoring**, in which as many as six major elements of writing are scored, thus enabling learners to home in on weaknesses and to capitalize on strengths.

Analytic scoring may be more appropriately called analytic *assessment* in order to capture its closer association with classroom language instruction than with formal testing. Brown and Bailey (1984) designed an analytical scoring scale that specified five major categories and a description of five different levels in each category, ranging from "unacceptable" to "excellent" (see Table 9.2).

At first glance, Brown and Bailey's scale may look similar to the TWE® holistic scale discussed earlier: for each scoring category there is a description that encompasses several subsets. A closer inspection, however, reveals much more detail in the analytic method. Instead of just six descriptions, there are 25, each subdivided into a number of contributing factors.

The order in which the five categories (organization, logical development of ideas, grammar, punctuation/spelling/mechanics, and style and quality of expression) are listed may bias the evaluator toward the greater importance of organization and logical development as opposed to punctuation and style. But the mathematical assignment of the 100-point scale gives equal weight (a maximum of 20 points) to each of the five major categories. Not all writing and assessment specialists agree. You might, for example, consider the analytical scoring profile suggested by Jacobs et al. (1981), in which five slightly different categories were given the point values shown on page 246.

2744

Table 9.2. Analytic scale for rating composition tasks (Brown & Bailey, 1984, pp. 39–41)

	20–18 Excellent to Good	17–15 Good to Adequate	14–12 Adequate to Fair	11–6 Unacceptable–not college-level work	5–1
I. Organization: Introduction, Body, and Conclusion	Appropriate title, effective introductory paragraph, topic is stated, leads to body; transitional expressions used; arrangement of material shows plan (could be outlined by reader); supporting evidence given for generalizations; conclusion logical and complete	Adequate title, introduction, and conclusion; body of essay is acceptable, but some evidence may be lacking, some ideas aren't fully developed; sequence is logical but transitional expressions may be absent or misused	Mediocre or scant introduction or conclusion; problems with the order of ideas in body; the generalizations may not be fully supported by the evidence given; problems of organization interfere	Shaky or minimally recognizable introduction; organization can barely be seen; severe problems with ordering of ideas; lack of supporting evidence; conclusion weak or illogical; inadequate effort at organization	Absence of introduction or conclusion; no apparent organization of body; severe lack of supporting evidence; writer has not made any effort to organize the composition (could not be outlined by reader)
II. Logical development of ideas: Content	Essay addresses the assigned topic; the ideas are concrete and thoroughly developed; no extraneous material; essay reflects thought	Essay addresses the issues but misses some points; ideas could be more fully developed; some extraneous material is present	Development of ideas not complete or essay is somewhat off the topic; paragraphs aren't divided exactly right	Ideas incomplete; essay does not reflect careful thinking or was hurriedly written; inadequate effort in area of content	Essay is completely inadequate and does not reflect college-level work; no apparent effort to consider the topic carefully

III. Grammar	Native-like fluency in English grammar; correct use of relative clauses, prepositions, modals, articles, verb forms, and tense sequencing; no fragments or run-on sentences	Advanced proficiency in English grammar; some grammar problems don't influence communication, although the reader is aware of them; no fragments or run-on sentences	Ideas are getting through to the reader, but grammar problems are apparent and have a negative effect on communication; run-on sentences or fragments present	Numerous serious grammar problems interfere with communication of the writer's ideas; grammar review of some areas clearly needed; difficult to read sentences	Severe grammar problems interfere greatly with the message; reader can't understand what the writer was trying to say; unintelligible sentence structure
IV. Punctuation, spelling, and mechanics	Correct use of English writing conventions: left and right margins, all needed capitals, paragraphs indented, punctuation and spelling; very neat	Some problems with writing conventions or punctuation; occasional spelling errors; left margin correct; paper is neat and legible	Uses general writing conventions but has errors; spelling problems distract reader; punctuation errors interfere with ideas	Serious problems with format of paper; parts of essay not legible; errors in sentence punctuation and final punctuation; unacceptable to educated readers	Complete disregard for English writing conventions; paper illegible; obvious capitals missing, no margins, severe spelling problems
V. Style and quality of expression	Precise vocabulary usage; use of parallel structures; concise; register good	Attempts variety; good vocabulary; not wordy; register OK; style fairly concise	Some vocabulary misused; lacks awareness of register; may be too wordy	Poor expression of ideas; problems in vocabulary; lacks variety of structure	Inappropriate use of vocabulary; no concept of register or sentence variety

276

Content	30
Organization	20
Vocabulary	20
Syntax	25
Mechanics	5
Total	**100**

As your curricular goals and students' needs vary, your own analytical scoring of essays may be appropriately tailored. Level of proficiency can make a significant difference in emphasis: at the intermediate level, for example, you might give more emphasis to syntax and mechanics, while advanced levels of writing may call for a strong push toward organization and development. Genre can also dictate variations in scoring. Would a summary of an article require the same relative emphases as a narrative essay? Most likely not. Certain types of writing, such as lab reports or interpretations of statistical data, may even need additional—or at least redefined—categories in order to capture the essential components of good writing within those genres.

Analytic scoring of compositions offers writers a little more washback than a single holistic or primary trait score. Scores in five or six major elements will help to call the writers' attention to areas of needed improvement. Practicality is lowered in that more time is required for teachers to attend to details within each of the categories in order to render a final score or grade, but ultimately students receive more information about their writing. Numerical scores alone, however, are still not sufficient for enabling students to become proficient writers, as we shall see in the next section.

BEYOND SCORING: RESPONDING TO EXTENSIVE WRITING

Formal testing carries with it the burden of designing a practical and reliable instrument that assesses its intended criterion accurately. To accomplish that mission, designers of writing tests are charged with the task of providing as "objective" a scoring procedure as possible, and one that in many cases can be easily interpreted by agents beyond the learner. Holistic, primary trait, and analytic scoring all satisfy those ends. Yet beyond mathematically calculated scores lies a rich domain of assessment in which a developing writer is coached from stage to stage in a process of building a storehouse of writing skills. Here in the classroom, in the tutored relationship of teacher and student, and in the community of peer learners, most of the hard work of assessing writing is carried out. Such assessment is informal, formative, and replete with washback.

Most writing specialists agree that the best way to teach writing is a hands-on approach that stimulates student output and then generates a series of self-assessments, peer editing and revision, and teacher response and conferencing (Raimes, 1991, 1998; Reid, 1993; Seow, 2002). It is not an approach that relies on a massive dose of lecturing

about good writing, nor on memorizing a bunch of rules about rhetorical organization, nor on sending students home with an assignment to turn in a paper the next day. People become good writers by writing and seeking the facilitative input of others to refine their skills.

Assessment takes on a crucial role in such an approach. Learning how to become a good writer places the student in an almost constant stage of assessment. To give the student the maximum benefit of assessment, it is important to consider (a) *earlier* stages (from freewriting to the first draft or two) and (b) *later* stages (revising and finalizing) of producing a written text. A further factor in assessing writing is the involvement of self, peers, and teacher at appropriate steps in the process. (For further guidelines on the process of teaching writing, see *TBP,* Chapter 19.)

Assessing Initial Stages of the Process of Composing

Following are some guidelines for assessing the initial stages (the first draft or two) of a written composition. These guidelines are generic for self, peer, and teacher responding. Each assessor will need to modify the list according to the level of the learner, the context, and the purpose in responding.

Assessment of initial stages in composing

1. Focus your efforts primarily on meaning, main idea, and organization.
2. Comment on the introductory paragraph.
3. Make general comments about the clarity of the main idea and logic or appropriateness of the organization.
4. As a rule of thumb, ignore minor (local) grammatical and lexical errors.
5. Indicate what appear to be major (global) errors (e.g., by underlining the text in question), but allow the writer to make corrections.
6. Do not rewrite questionable, ungrammatical, or awkward sentences; rather, probe with a question about meaning.
7. Comment on features that appear to be irrelevant to the topic.

The teacher-assessor's role is as a guide, a facilitator, and an ally; therefore, assessment at this stage of writing needs to be as positive as possible to encourage the writer. An early focus on overall structure and meaning will enable writers to clarify their purpose and plan and will set a framework for the writers' later refinement of the lexical and grammatical issues.

Assessing Later Stages of the Process of Composing

Once the writer has determined and clarified his or her purpose and plan, and has completed at least one or perhaps two drafts, the focus shifts toward "fine tuning" the expression with a view toward a final revision. Editing and responding assume an appropriately different character now, with these guidelines:

278

Assessment of later stages in composing

---

1. Comment on the specific clarity and strength of all main ideas and supporting ideas, and on argument and logic.
2. Call attention to minor ("local") grammatical and mechanical (spelling, punctuation) errors, but direct the writer to self-correct.
3. Comment on any further word choices and expressions that may not be awkward but are not as clear or direct as they could be.
4. Point out any problems with cohesive devices within and across paragraphs.
5. If appropriate, comment on documentation, citation of sources, evidence, and other support.
6. Comment on the adequacy and strength of the conclusion.

Through all these stages it is assumed that peers and teacher are both responding to the writer through conferencing in person, electronic communication, or, at the very least, an exchange of papers. The impromptu timed tests and the methods of scoring discussed earlier may appear to be only distantly related to such an individualized process of creating a written text, but are they, in reality? All those developmental stages may be the preparation that learners need both to function in creative real-world writing tasks and to successfully demonstrate their competence on a timed impromptu test. And those holistic scores are after all generalizations of the various components of effective writing. If the hard work of successfully progressing through a semester or two of a challenging course in academic writing ultimately means that writers are ready to function in their real-world contexts, *and* to get a 5 or 6 on the TWE, then all the effort was worthwhile.

§   §   §   §   §

This chapter completes the cycle of considering the assessment of all of the four skills of listening, speaking, reading, and writing. As you contemplate using some of the assessment techniques that have been suggested, I think you can now fully appreciate two significant overarching guidelines for designing an effective assessment procedure:

1. It is virtually impossible to isolate any one of the four skills without the involvement of at least one other mode of performance. Don't underestimate the power of the integration of skills in assessments designed to target a single skill area.

2. The variety of assessment techniques and item types and tasks is virtually infinite in that there is always some possibility for creating a unique variation. Explore those alternatives, but with some caution lest your overzealous urge to be innovative distract you from a central focus on achieving the intended purpose and rendering an appropriate evaluation of performance.

279

EXERCISES

[Note: (I) Individual work; (G) Group or pair work; (C) Whole-class discussion.]

1. (C) Genres of reading were listed at the beginning of Chapter 8, and genres of writing in this chapter, a shorter list. Why is the list for writing shorter? Add other examples to each of the three categories. Among the listed examples and new ones you come up with, be specific in citing what makes some genres more difficult than others. Select a few of the more difficult genres and discuss what you would assess (criteria) and how you would assess (some possible assessment techniques) them.

2. (C) Review the four basic types of writing that were outlined at the beginning of the chapter. Offer examples of each and pay special attention to distinguishing between imitative and intensive, and between responsive and extensive.

3. (G) Look at the list of micro- and macroskills of writing on page 221. In pairs, each assigned to a different skill (or two), brainstorm some tasks that assess those skills. Present your findings to the rest of the class.

4. (C) In Chapter 6, eight characteristics of listening were listed (page 122) that make listening "difficult." What makes writing difficult? Devise a similar list, which could form a set of specifications to pay special attention to in assessing writing.

5. (G) Divide the four basic types of writing among groups or pairs, one type for each. Look at the sample assessment techniques provided and evaluate them according to the five principles (practicality, reliability, validity [especially face and content], authenticity, and washback). Present your critique to the rest of the class.

6. (G) In the same groups as #5 above with the same type of writing, design some other item types, different from the one(s) provided here, that assess the same type of writing performance.

7. (I/C) Visit the TOEFL website and click on the description of the Test of Written English to familiarize yourself further with the TWE. Then, do the following: (a) Look at the TWE holistic scoring guide (page 239) and evaluate its rater reliability. (b) Discuss the validity of a timed impromptu test such as this for admission to an English-speaking university.

8. (C) Review the advantages and disadvantages of the three kinds of scoring presented in this chapter: holistic, primary trait, and analytic. Construct a chart that shows how different contexts (types of test, objectives of a curriculum, proficiency levels, etc.) may benefit from each kind of scoring.

280