TESTING

READING COMPREHENSION

CHAPTER 8

ASSESSING READING

Even as we are bombarded with an unending supply of visual and auditory media, the written word continues in its function to convey information, to amuse and entertain us, to codify our social, economic, and legal conventions, and to fulfill a host of other functions. In literate societies, most "normal" children learn to read by the age of five or six, and some even earlier. With the exception of a small number of people with learning disabilities, reading is a skill that is taken for granted.

In foreign language learning, reading is likewise a skill that teachers simply expect learners to acquire. Basic, beginning-level textbooks in a foreign language presuppose a student's reading ability if only because it's a *book* that is the medium. Most formal tests use the written word as a stimulus for test-taker response; even oral interviews may require reading performance for certain tasks. Reading, arguably the most essential skill for success in all educational contexts, remains a skill of paramount importance as we create assessments of general language ability.

Is reading so natural and normal that learners should simply be exposed to written texts with no particular instruction? Will they just absorb the skills necessary to convert their perception of a handful of letters into meaningful chunks of information? Not necessarily. For learners of English, two primary hurdles must be cleared in order to become efficient readers. First, they need to be able to master fundamental **bottom-up** strategies for processing separate letters, words, and phrases, as well as **top-down,** conceptually driven strategies for comprehension. Second, as part of that top-down approach, second language readers must develop appropriate content and formal **schemata**—background information and cultural experience—to carry out those interpretations effectively.

The assessment of reading ability does not end with the measurement of comprehension. Strategic pathways to full understanding are often important factors to include in assessing learners, especially in the case of most classroom assessments that are **formative** in nature. An inability to comprehend may thus be traced to a need to enhance a test-taker's strategies for achieving ultimate comprehension. For example, an academic technical report may be comprehensible to a student at the sentence level, but if the learner has not exercised certain strategies for noting the discourse conventions of that genre, misunderstanding may occur.

As we consider a number of different types or genres of written texts, the components of reading ability, and specific tasks that are commonly used in the assessment of reading, let's not forget the unobservable nature of reading. Like listening, one cannot see the **process** of reading, nor can one observe a specific **product** of reading. Other than observing a reader's eye movements and page turning, there is no technology that enables us to "see" sequences of graphic symbols traveling from the pages of a book into compartments of the brain (in a possible bottom-up process). Even more outlandish is the notion that one might be able to watch information from the brain make its way down onto the page (in typical top-down strategies). Further, once something is read—information from the written text is stored—no technology allows us to empirically measure exactly what is lodged in the brain. All assessment of reading must be carried out by inference.

TYPES (GENRES) OF READING

Each type or **genre** of written text has its own set of governing rules and conventions. A reader must be able to anticipate those conventions in order to process meaning efficiently. With an extraordinary number of genres present in any literate culture, the reader's ability to process texts must be very sophisticated. Consider the following abridged list of common genres, which ultimately form part of the specifications for assessments of reading ability.

Genres of reading

1. **Academic reading**
   - general interest articles (in magazines, newspapers, etc.)
   - technical reports (e.g., lab reports), professional journal articles
   - reference material (dictionaries, etc.)
   - textbooks, theses
   - essays, papers
   - test directions
   - editorials and opinion writing
2. **Job-related reading**
   - messages (e.g., phone messages)
   - letters/emails
   - memos (e.g., interoffice)
   - reports (e.g., job evaluations, project reports)
   - schedules, labels, signs, announcements
   - forms, applications, questionnaires
   - financial documents (bills, invoices, etc.)
   - directories (telephone, office, etc.)
   - manuals, directions

3. **Personal reading**
   newspapers and magazines
   letters, emails, greeting cards, invitations
   messages, notes, lists
   schedules (train, bus, plane, etc.)
   recipes, menus, maps, calendars
   advertisements (commercials, want ads)
   novels, short stories, jokes, drama, poetry
   financial documents (e.g., checks, tax forms, loan applications)
   forms, questionnaires, medical reports, immigration documents
   comic strips, cartoons

When we realize that this list is only the beginning, it is easy to see how overwhelming it is to learn to read in a foreign language! The genre of a text enables readers to apply certain **schemata** that will assist them in extracting appropriate meaning. If, for example, readers know that a text is a recipe, they will expect a certain arrangement of information (ingredients) and will know to search for a sequential order of directions. Efficient readers also have to know what their purpose is in reading a text, the strategies for accomplishing that purpose, and how to retain the information.

The content validity of an assessment procedure is largely established through the genre of a text. For example, if learners in a program of English for tourism have been learning how to deal with customers needing to arrange bus tours, then assessments of their ability should include guidebooks, maps, transportation schedules, calendars, and other relevant texts.

MICROSKILLS, MACROSKILLS, AND STRATEGIES FOR READING

Aside from attending to genres of text, the skills and strategies for accomplishing reading emerge as a crucial consideration in the assessment of reading ability. The micro- and macroskills below represent the spectrum of possibilities for objectives in the assessment of reading comprehension.

Micro- and macroskills for reading comprehension

Microskills

1. Discriminate among the distinctive graphemes and orthographic patterns of English.
2. Retain chunks of language of different lengths in short-term memory.
3. Process writing at an efficient rate of speed to suit the purpose.

4. Recognize a core of words, and interpret word order patterns and their significance.
5. Recognize grammatical word classes (nouns, verbs, etc.), systems (e.g., tense, agreement, pluralization), patterns, rules, and elliptical forms.
6. Recognize that a particular meaning may be expressed in different grammatical forms.
7. Recognize cohesive devices in written discourse and their role in signaling the relationship between and among clauses.

Macroskills

8. Recognize the rhetorical forms of written discourse and their significance for interpretation.
9. Recognize the communicative functions of written texts, according to form and purpose.
10. Infer context that is not explicit by using background knowledge.
11. From described events, ideas, etc., infer links and connections between events, deduce causes and effects, and detect such relations as main idea, supporting idea, new information, given information, generalization, and exemplification.
12. Distinguish between literal and implied meanings.
13. Detect culturally specific references and interpret them in a context of the appropriate cultural schemata.
14. Develop and use a battery of reading strategies, such as scanning and skimming, detecting discourse markers, guessing the meaning of words from context, and activating schemata for the interpretation of texts.

The assessment of reading can imply the assessment of a storehouse of reading strategies, as indicated in item #14. Aside from simply testing the ultimate achievement of comprehension of a written text, it may be important in some contexts to assess one or more of a storehouse of classic reading strategies. The brief taxonomy of strategies below is a list of possible assessment criteria.

Some principal strategies for reading comprehension

1. Identify your purpose in reading a text.
2. Apply spelling rules and conventions for bottom-up decoding.
3. Use lexical analysis (prefixes, roots, suffixes, etc.) to determine meaning.
4. Guess at meaning (of words, idioms, etc.) when you aren't certain.
5. Skim the text for the gist and for main ideas.
6. Scan the text for specific information (names, dates, key words).
7. Use silent reading techniques for rapid processing.

Written Response

The same stimuli are presented, and the test-taker's task is to reproduce the probe in writing. Because of the transfer across different skills here, evaluation of the test-taker's response must be carefully treated. If an error occurs, make sure you determine its source; what might be assumed to be a writing error, for example, may actually be a reading error, and vice versa.

Multiple-Choice

Multiple-choice responses are not only a matter of choosing one of four or five possible answers. Other formats, some of which are especially useful at the low levels of reading, include same/different, circle the answer, true/false, choose the letter, and matching. Here are some possibilities.

Minimal pair distinction

*Test-takers read:** Circle "S" for same or "D" for different.

1.	led	let	S	D
2.	bit	bit	S	D
3.	seat	set	S	D
4.	too	to	S	D

**In the case of very low level learners, the teacher/administrator reads directions.*

Grapheme recognition task

*Test-takers read:** Circle the "odd" item, the one that doesn't "belong."

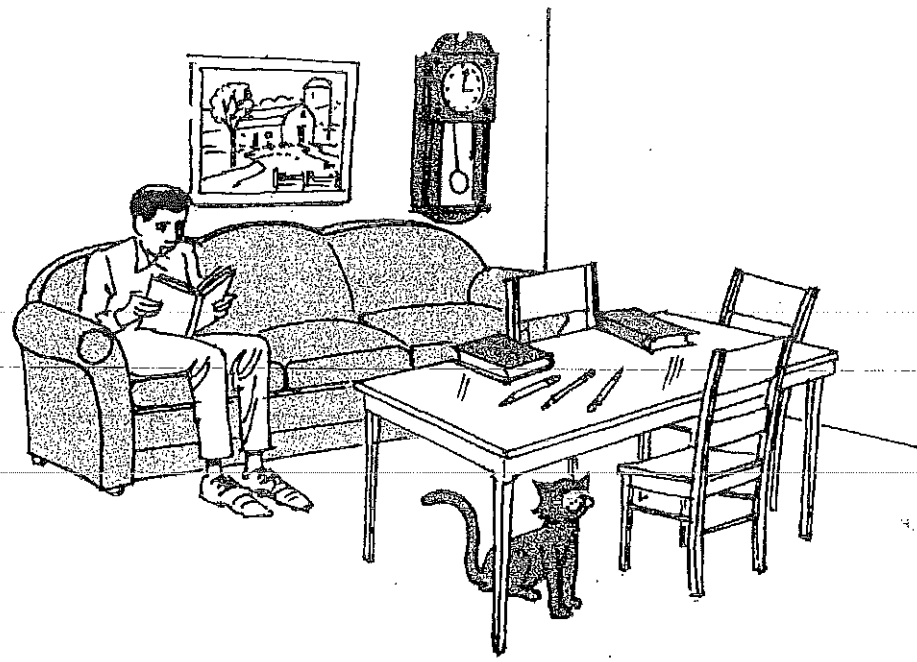
1.	piece	peace	piece
2.	book	book	boot

**In the case of very low level learners, the teacher/administrator reads directions.*

Picture-Cued Items

Test-takers are shown a picture, such as the one on the next page, along with a written text and are given one of a number of possible tasks to perform.

Picture-cued word identification (Brown & Sahni, 1994, p. 124)



Test-takers hear: Point to the word that you read here.

cat	clock	chair

With the same picture, the test-taker might read sentences and then point to the correct part of the picture:

Picture-cued sentence identification

Test-takers hear: Point to the part of the picture that you read about here.

Test-takers see the picture and read each sentence written on a separate card.

The man is reading a book.

The cat is under the table.

8. Use marginal notes, outlines, charts, or semantic maps for understanding and retaining information.
9. Distinguish between literal and implied meanings.
10. Capitalize on discourse markers to process relationships.

TYPES OF READING

In the previous chapters we saw that both listening and speaking could be subdivided into at least five different types of listening and speaking performance. In the case of reading, variety of performance is derived more from the multiplicity of types of texts (the genres listed above) than from the variety of overt types of performance. Nevertheless, for considering assessment procedures, several types of reading performance are typically identified, and these will serve as organizers of various assessment tasks.

1. Perceptive. In keeping with the set of categories specified for listening comprehension, similar specifications are offered here, except with some differing terminology to capture the uniqueness of reading. Perceptive reading tasks involve attending to the *components* of larger stretches of discourse: letters, words, punctuation, and other graphemic symbols. Bottom-up processing is implied.

2. Selective. This category is largely an artifact of assessment formats. In order to ascertain one's reading recognition of lexical, grammatical, or discourse features of language within a very short stretch of language, certain typical tasks are used: picture-cued tasks, matching, true/false, multiple-choice, etc. Stimuli include sentences, brief paragraphs, and simple charts and graphs. Brief responses are intended as well. A combination of bottom-up and top-down processing may be used.

3. Interactive. Included among interactive reading types are stretches of language of several paragraphs to one page or more in which the reader must, in a psycholinguistic sense, *interact* with the text. That is, reading is a process of negotiating meaning; the reader brings to the text a set of schemata for understanding it, and intake is the product of that interaction. Typical genres that lend themselves to interactive reading are anecdotes, short narratives and descriptions, excerpts from longer texts, questionnaires, memos, announcements, directions, recipes, and the like. The focus of an interactive task is to identify relevant features (lexical, symbolic, grammatical, and discourse) within texts of moderately short length with the objective of retaining the information that is processed. Top-down processing is typical of such tasks, although some instances of bottom-up performance may be necessary.

4. Extensive. Extensive reading, as discussed in this book, applies to texts of more than a page, up to and including professional articles, essays, technical reports, short stories, and books. (It should be noted that reading research commonly refers to "extensive reading" as longer stretches of discourse, such as long articles and books that are usually read outside a classroom hour. Here that definition is

massaged a little in order to encompass any text longer than a page.) The purposes of assessment usually are to tap into a learner's global understanding of a text, as opposed to asking test-takers to "zoom in" on small details. Top-down processing is assumed for most extensive tasks.

The four types of reading are demonstrated in Figure 8.1, which shows the relationships of length, focus, and processing mode among the four types.

	Length			Focus		Process	
	Short	Medium	Long	Form	Meaning	Bottom-Up	Top-Down
Perceptive	••			••		••	
Selective	•	•		••	•	•	•
Interactive		••		•	••	•	••
Extensive			••		••		••

•• strong emphasis
• moderate emphasis

Figure 8.1. Types of reading by length, focus, and process

DESIGNING ASSESSMENT TASKS: PERCEPTIVE READING

At the beginning level of reading a second language lies a set of tasks that are fundamental and basic: recognition of alphabetic symbols, capitalized and lowercase letters, punctuation, words, and grapheme-phoneme correspondences. Such tasks of perception are often referred to as **literacy** tasks, implying that the learner is in the early stages of becoming "literate." Some learners are already literate in their own native language, but in other cases the second language may be the first language that they have ever learned to read. This latter context poses cognitive and sometimes age-related issues that need to be considered carefully. Assessment of literacy is no easy assignment, and if you are interested in this particular challenging area, further reading beyond this book is advised (Harp, 1991; Farr & Tone, 1994; Genesee, 1994; Cooper, 1997). Assessment of basic reading skills may be carried out in a number of different ways.

Reading Aloud

The test-taker sees separate letters, words, and/or short sentences and reads them aloud, one by one, in the presence of an administrator. Since the assessment is of *reading* comprehension, any recognizable oral approximation of the target response is considered correct.

Or a true/false procedure might be presented with the same picture cue:

Picture-cued true/false sentence identification

Test-takers read:

1. The pencils are under the table. T F
2. The cat is on the table. T F
3. The picture is over the couch. T F

Matching can be an effective method of assessing reading at this level. With objects labeled A, B, C, D, E in the picture, the test-taker reads words and writes the appropriate letter beside the word:

Picture-cued matching word identification

Test-takers read:

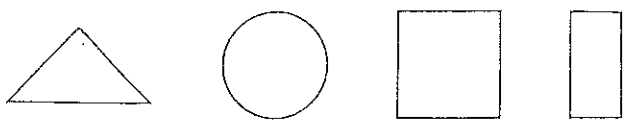
1. clock _____
2. chair _____
3. books _____
4. cat _____
5. table _____

Finally, test-takers might see a word or phrase and then be directed to choose one of four pictures that is being described, thus requiring the test-taker to transfer from a verbal to a nonverbal mode. In the following item, test-takers choose the correct letter:

Multiple-choice picture-cued word identification

Test-takers read: Rectangle

Test-takers see, and choose the correct item:



DESIGNING ASSESSMENT TASKS: SELECTIVE READING

Just above the rudimentary skill level of perception of letters and words is a category in which the test designer focuses on formal aspects of language (lexical, grammatical, and a few discourse features). This category includes what many incorrectly think of as testing "vocabulary and grammar." How many textbooks provide little tests and quizzes labeled "vocabulary and grammar" and never feature any other skill besides reading? Lexical and grammatical aspects of language are simply the forms we use to perform all four of the skills of listening, speaking, reading, and writing. (Notice that in all of these chapters on the four skills, formal features of language have become a potential focus for assessment.)

Here are some of the possible tasks you can use to assess lexical and grammatical aspects of *reading* ability.

Multiple-Choice (for Form-Focused Criteria)

By far the most popular method of testing a reading knowledge of vocabulary and grammar is the multiple-choice format, mainly for reasons of practicality: it is easy to administer and can be scored quickly. The most straightforward multiple-choice items may have little context, but might serve as a vocabulary or grammar check.

Multiple-choice vocabulary/grammar tasks

1. He's not married. He's _____.
   A. young
   B. single
   C. first
   D. a husband
2. If there's no doorbell, please _____ on the door.
   A. kneel
   B. type
   C. knock
   D. shout
3. The mouse is _____ the bed.
   A. under
   B. around
   C. between
4. The bank robbery occurred _____ I was in the restroom.
   A. that
   B. during
   C. while
   D. which

5. Yeast is an organic catalyst _____ known to prehistoric humanity.
   A. was
   B. which was
   C. which it
   D. which

This kind of darting from one context to another to another in a test has become so commonplace that learners almost expect the disjointedness. Some improvement of these items is possible by providing some context within each item:

Contextualized multiple-choice vocabulary/grammar tasks

1. Oscar: Do you like champagne?
   Lucy: No, I can't _____ it!
   A. stand
   B. prefer
   C. hate
2. Manager: Do you like to work by yourself?
   Employee: Yes, I like to work _____.
   A. independently
   B. definitely
   C. impatiently
3. Jack: Do you have a coat like this?
   John: Yes, mine is _____ yours.
   A. so same as
   B. the same like
   C. as same as
   D. the same as
4. Boss: Where did I put the Johnson file?
   Sectretary: I think _____ is on your desk.
   A. you were the file looking at
   B. the you were looking at file
   C. the file you were looking at
   D. you were looking at the file

A better contextualized format is to offer a modified cloze test (see page 201 for a treatment of cloze testing) adjusted to fit the objectives being assessed. In the example below, a few lines of English add to overall context.

Multiple-choice cloze vocabulary/grammar task

I've lived in the United States (**21**) _____ three years. I (**22**) _____ live in Costa Rica. I (**23**) _____ speak any English. I used to (**24**) _____ homesick, but now I enjoy (**25**) _____ here. I have never (**26**) _____ back home (**27**) _____ I came to the United States, but I might (**28**) _____ to visit my family soon.

21. A. since
B. for
C. during

22. A. used to
B. use to
C. was

23. A. couldn't
B. could
C. can

24. A. been
B. be
C. being

25. A. live
B. to live
C. living

26. A. be
B. been
C. was

27. A. when
B. while
C. since

28. A. go
B. will go
C. going

The context of the story in this example may not specifically help the test-taker to respond to the items more easily, but it allows the learner to attend to one set of related sentences for eight items that assess vocabulary and grammar. Other contexts might involve some content dependencies, such that earlier sentences predict the correct response for a later item. Thus, a pair of sentences in a short narrative might read:

He showed his suitcase (**29**) _____ me, but it wasn't big (**30**) _____ to fit all his clothes. So I gave him my suitcase, which was (**31**) _____.

29. A. for
B. from
C. to

30. A. so
B. too
C. enough

31. A. larger
B. smaller
C. largest

To respond to item #31 correctly, the test-taker needs to be able to comprehend the context of needing a *larger,* but not an equally grammatically correct *smaller,* suitcase. While such dependencies offer greater authenticity to an assessment, they also add the potential problem of a test-taker's missing several later items because of an earlier comprehension error.

Matching Tasks

At this selective level of reading, the test-taker's task is simply to respond correctly, which makes matching an appropriate format. The most frequently appearing criterion in matching procedures is vocabulary. Following is a typical format:

Vocabulary matching task

Write in the letter of the definition on the right that matches the word on the left.

_____ **1.** exhausted	a. unhappy
_____ **2.** disappointed	b. understanding of others
_____ **3.** enthusiastic	c. tired
_____ **4.** empathetic	d. excited

To add a communicative quality to matching, the first numbered list is sometimes a set of sentences with blanks in them, with a list of words to choose from:

Selected response fill-in vocabulary task

1. At the end of the long race, the runners were totally _____.
2. My parents were _____ with my bad performance on the final exam.
3. Everyone in the office was _____ about the new salary raises.
4. The _____ listening of the counselor made Christina feel well understood.

Choose from among the following:

disappointed
empathetic
exhausted
enthusiastic

Alderson (2000, p. 218) suggested matching procedures at an even more sophisticated level, where test-takers have to discern pragmatic interpretations of certain signs or labels such as "Freshly made sandwiches" and "Use before 10/23/02." Matches for those two are "We sell food" and "This is too old," which are selected from a number of other options.

Matching tasks have the advantage of offering an alternative to traditional multiple-choice or fill-in-the-blank formats and are sometimes easier to construct than multiple-choice items, as long as the test designer has chosen the matches carefully. Some disadvantages do come with this framework, however. They can become more of a puzzle-solving process than a genuine test of comprehension as test-takers struggle with the search for a match, possibly among 10 or 20 different items. Like other tasks in this section, they also are contrived exercises that are endemic to academia that will seldom be found in the real world.

Editing Tasks

Editing for grammatical or rhetorical errors is a widely used test method for assessing linguistic competence in reading. The TOEFL® and many other tests employ this technique with the argument that it not only focuses on grammar but also introduces a simulation of the authentic task of editing, or discerning errors in written passages. Its authenticity may be supported if you consider proofreading as a real-world skill that is being tested. Here is a typical set of examples of editing.

Multiple-choice grammar editing task (Phillips, 2001, p. 219)

Test-takers read: Choose the letter of the underlined word that is not correct.

1. The abrasively action of the wind wears away softer layers of rock.
   A B C D

2. There are two way of making a gas condense: cooling it or putting it under
   A B C D
   pressure.

3. Researchers have discovered that the application of bright light can sometimes
   A B
   be uses to overcome jet lag.
   C D

The above examples, with their disparate subject-matter content, are not as authentic as asking test-takers to edit a whole essay (see discussion below, pages 207–208). Of course, if learners have never practiced error detection tasks, the task itself is of some difficulty. Nevertheless, error detection has been

shown to be positively correlated with both listening comprehension and reading comprehension results on the TOEFL, at r = .58 and .76, respectively (*TOEFL Score User Guide,* 2001). Despite some authenticity quibbles, this task maintains a construct validity that justifies its use.

Picture-Cued Tasks

In the previous section we looked at picture-cued tasks for perceptive recognition of symbols and words. Pictures and photographs may be equally well utilized for examining ability at the selective level. Several types of picture-cued methods are commonly used.

1. Test-takers read a sentence or passage and choose one of four pictures that is being described. The sentence (or sentences) at this level is more complex. A computer-based example follows:

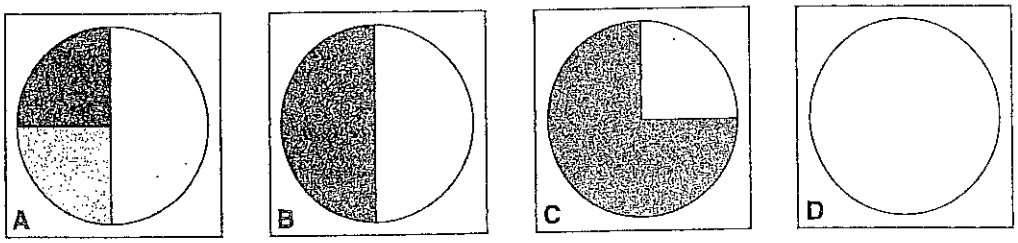
Multiple-choice picture-cued response (Phillips, 2001, p. 276)

Test-takers read a three-paragraph passage, one sentence of which is:

During at least three quarters of the year, the Arctic is frozen.

Click on the chart that shows the relative amount of time each year that water is available to plants in the Arctic.

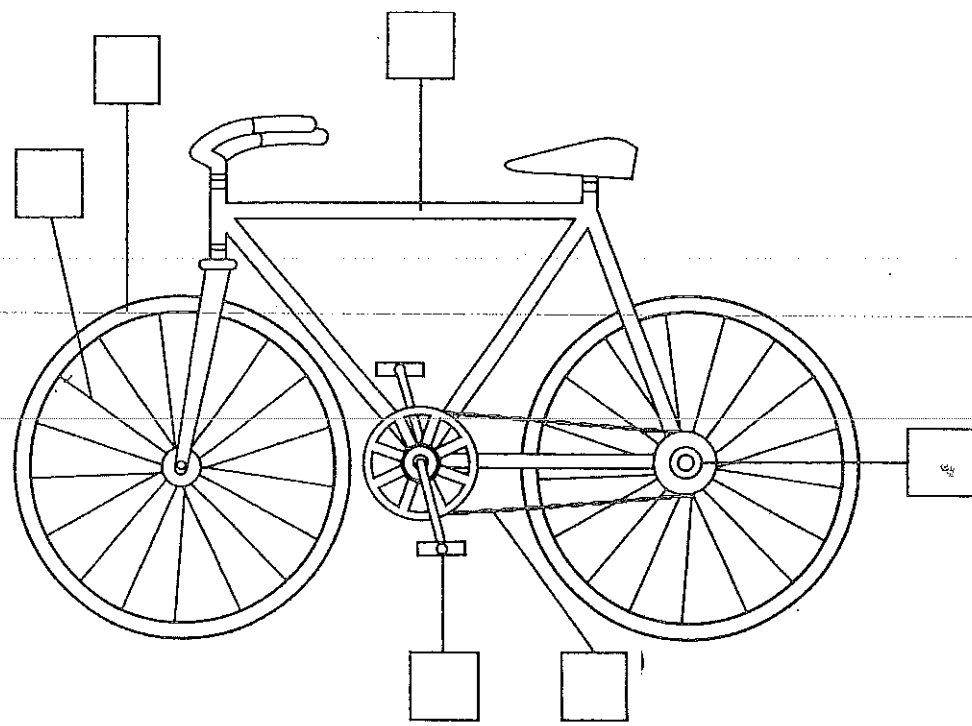
Test-takers see the following four pictures:



2. Test-takers read a series of sentences or definitions, each describing a labeled part of a picture or diagram. Their task is to identify each labeled item. In the following diagram, test-takers do not necessarily know each term, but by reading the definition they are able to make an identification. For example:

Diagram-labeling task

Test-takers see:



Test-takers read:

Label the picture with the number of the corresponding item described below.

1. wire supports extending from the hub of a wheel to its perimeter
2. a long, narrow support pole between the seat and the handlebars
3. a small, geared wheel concentric with the rear wheel
4. a long, linked, flexible metal device that propels the vehicle
5. a small rectangular lever operated by the foot to propel the vehicle
6. a tough but somewhat flexible rubber item that circles each wheel

The essential difference between the picture-cued tasks here and those that were outlined in the previous section is the complexity of the language.

Gap-Filling Tasks

Many of the multiple-choice tasks described above can be converted into gap-filling, or "fill-in-the-blank," items in which the test-taker's response is to write a word or phrase. An extension of simple gap-filling tasks is to create sentence completion items where test-takers read part of a sentence and then complete it by writing a phrase.

Sentence completion tasks

Oscar:	Doctor, what should I do if I get sick?
Doctor:	It is best to stay home and _____.
	If you have a fever, _____.
	You should drink as much _____.
	The worst thing you can do is _____.
	You should also _____.

The obvious disadvantage of this type of task is its questionable assessment of reading ability. The task requires both reading and writing performance, thereby rendering it of low validity in isolating reading as the sole criterion. Another drawback is scoring the variety of creative responses that are likely to appear. You will have to make a number of judgment calls on what comprises a correct response. In a test of reading comprehension only, you must accept as correct any responses that demonstrate comprehension of the first part of the sentence. This alone indicates that such tasks are better categorized as integrative procedures.

DESIGNING ASSESSMENT TASKS: INTERACTIVE READING

Tasks at this level, like selective tasks, have a combination of form-focused and meaning-focused objectives but with more emphasis on meaning. Interactive tasks may therefore imply a little more focus on top-down processing than on bottom-up. Texts are a little longer, from a paragraph to as much as a page or so in the case of ordinary prose. Charts, graphs, and other graphics may be somewhat complex in their format.

Cloze Tasks

One of the most popular types of reading assessment task is the **cloze** procedure. The word *cloze* was coined by educational psychologists to capture the Gestalt psychological concept of "closure," that is, the ability to fill in gaps in an incomplete image (visual, auditory, or cognitive) and supply (from background schemata) omitted details.

In written language, a sentence with a word left out should have enough context that a reader can close that gap with a calculated guess, using linguistic expectancies (formal schemata), background experience (content schemata), and some strategic competence. Based on this assumption, cloze tests were developed for native language readers and defended as an appropriate gauge of reading ability. Some research (Oller, 1973, 1976, 1979) on second language acquisition vigorously defends cloze testing as an integrative measure not only of reading ability but also

of other language abilities. It was argued that the ability to make coherent guesses in cloze gaps also taps into the ability to listen, speak, and write. With the decline of zeal for the search for the ideal integrative test in recent years, cloze testing has returned to a more appropriate status as one of a number of assessment procedures available for testing reading ability.

Cloze tests are usually a minimum of two paragraphs in length in order to account for discourse expectancies. They can be constructed relatively easily as long as the specifications for choosing deletions and for scoring are clearly defined. Typically every seventh word (plus or minus two) is deleted (known as **fixed-ratio deletion**), but many cloze test designers instead use a **rational deletion** procedure of choosing deletions according to the grammatical or discourse functions of the words. Rational deletion also allows the designer to avoid deleting words that would be difficult to predict from the context. For example, in the sentence "Everyone in the crowd enjoyed the gorgeous sunset," the seventh word is *gorgeous*, but learners could easily substitute other appropriate adjectives. Traditionally, cloze passages have between 30 and 50 blanks to fill, but a passage with as few as half a dozen blanks can legitimately be labeled a cloze test.

Two approaches to the scoring of cloze tests are commonly used. The **exact word** method gives credit to test-takers only if they insert the exact word that was originally deleted. The second method, **appropriate word** scoring, credits the test-taker for supplying any word that is grammatically correct and that makes good sense in the context. In the sentence above about the "gorgeous sunset," the test-takers would get credit for supplying *beautiful, amazing,* and *spectacular.* The choice between the two methods of scoring is one of practicality/reliability vs. face validity. In the exact word approach, scoring can be done quickly (especially if the procedure uses a multiple-choice technique) and reliably. The second approach takes more time because the teacher must determine whether each response is indeed appropriate, but students will perceive the test as being fairer: they won't get "marked off" for appropriate, grammatically correct responses.

The following excerpts from a longer essay illustrate the difference between rational and fixed-ratio deletion, and between exact word and appropriate word scoring.

Cloze procedure, fixed-ratio deletion (every seventh word)

> The recognition that one's feelings of (1) _____ and unhappiness can coexist much like (2) _____ and hate in a close relationship (3) _____ offer valuable clues on how to (4) _____ a happier life. It suggests, for (5) _____, that changing or avoiding things that (6) _____ you miserable may well make you (7) _____ miserable but probably no happier.

Cloze procedure, rational deletion (prepositions and conjunctions)

> The recognition that one's feelings (1) _____ happiness (2) _____ unhappiness can coexist much like love and hate (3) _____ a close relationship may offer valuable clues (4) _____ how to lead a happier life. It suggests, (5) _____ example, that changing (6) _____ avoiding things that make you miserable may well make you less miserable (7) _____ probably no happier.

In both versions there are seven deletions, but the second version allows the test designer to tap into prediction of prepositions and conjunctions in particular. And the second version provides more washback as students focus on targeted grammatical features.

Both of the scoring methods named above could present problems, with the first version presenting a little more ambiguity. Possible responses might include:

Fixed-ratio version, blank #3: *may, might, could, can*
#4: *lead, live, have, seek*
#5: *example, instance*

Rational deletion version, blank #4: *on, about*
#6: *or, and*
#7: *but, and*

Arranging a cloze test in a multiple-choice format allows even more rapid scoring: hand scoring with an answer key or hole-punched grid, or computer scoring using scannable answer sheets. Multiple-choice cloze tests must of course adhere to all the other guidelines for effective multiple-choice items that were covered in Chapter 4, especially the choice of appropriate distractors; therefore they can take much longer to construct—possibly too long to pay off in a classroom setting.

Some variations on standard cloze testing have appeared over the years; two of the better known are the C-test and the cloze-elide procedure. In the C-test (Klein-Braley & Raatz, 1984; Klein-Braley, 1985; Dörnyei & Katona, 1992), the second half (according to the number of letters) of every other word is obliterated and the test-taker must restore each word. While Klein-Braley and others vouched for its validity and reliability, many consider this technique to be "even more irritating to complete than cloze tests" (Alderson, 2000, p. 225). Look at the following example and judge for yourself:

C-test procedure

> The recognition th_ _ one's feel_ _ _ _ of happ_ _ _ _ _ and unhap_ _ _ _ _ _ can coe_ _ _ _ much li_ _ love a_ _ hate i_ a cl_ _ _ relati_ _ _ _ _ _ may of_ _ _ valuable cl_ _ _ on h_ _ to le_ _ a hap_ _ _ _ life. I_ suggests, f_ _ example, th_ _ changing o_ avoiding thi_ _ _ that ma_ _ you mise_ _ _ _ _ may we_ _ make y_ _ less mise_ _ _ _ _ but prob_ _ _ _ no hap_ _ _ _.

The second variation, the **cloze-elide** procedure, inserts words into a text that don't belong. The test-taker's task is to detect and cross out the "intrusive" words. Look at the same familiar passage:

Cloze-elide procedure

> The recognition that one's now feelings of happiness and unhappiness can under coexist much like love and hate in a close then relationship may offer valuable clues on how to lead a happier with life. It suggests, for example, that changing or avoiding my things that make you miserable may well make you less miserable ever but probably no happier.

Critics of this procedure (Davies, 1975) claimed that the cloze-elide procedure is actually a test of reading speed and not of proofreading skill, as its proponents asserted. Two disadvantages are nevertheless immediately apparent: (1) Neither the words to insert nor the frequency of insertion appears to have any rationale. (2) Fast and efficient readers are not adept at detecting the intrusive words. Good readers naturally weed out such potential interruptions.

Impromptu Reading Plus Comprehension Questions

If cloze testing is the most-researched procedure for assessing reading, the traditional "Read a passage and answer some questions" technique is undoubtedly the oldest and the most common. Virtually every proficiency test uses the format, and one would rarely consider assessing reading without some component of the assessment involving impromptu reading and responding to questions.

In Chapter 4, in the discussion on proficiency testing, we looked at a typical reading comprehension passage and a set of questions from the TOEFL. Here's another such passage.

Reading comprehension passage (Phillips, 2001, pp. 421–422)

Questions 1–10

The Hollywood sign in the hills that line the northern border of Los Angeles is a famous landmark recognized the world over. The white-painted, 50-foot-high, sheet metal letters can be seen from great distances across the Los Angeles basin.

Line (5) The sign was not constructed, as one might suppose, by the movie business as a means of celebrating the importance of Hollywood to this industry; instead, it was first constructed in 1923 as a means of advertising homes for sale in a 500-acre housing subdivision in a part of Los Angeles called "Hollywoodland." The sign that was constructed at the time, of course, said "Hollywoodland." Over the years, people began referring to the area by the shortened version "Hollywood," and after the sign and its site were donated to the city in 1945, the last four letters were removed.

(10) The sign suffered from years of disrepair, and in 1973 it needed to be completely replaced, at a cost of $27,700 per letter. Various celebrities were instrumental in helping to raise needed funds. Rock star Alice Cooper, for example, bought an O in memory of Groucho Marx, and Hugh Hefner of *Playboy* fame held a benefit party to raise the money for the Y. The construction of the new sign was finally completed in 1978.

1. What is the topic of this passage?
   (A) A famous sign
   (B) A famous city
   (C) World landmarks
   (D) Hollywood versus Hollywoodland
2. The expression "the world over" in line 2 could best be replaced by
   (A) in the northern parts of the world
   (B) on top of the world
   (C) in the entire world
   (D) in the skies
3. It can be inferred from the passage that most people think that the Hollywood sign was first constructed by
   (A) an advertising company
   (B) the movie industry
   (C) a construction company
   (D) the city of Los Angeles
4. The pronoun "it" in line 5 refers to
   (A) the sign
   (B) the movie business
   (C) the importance of Hollywood
   (D) this industry
5. According to the passage, the Hollywood sign was first built in
   (A) 1923
   (B) 1949
   (C) 1973
   (D) 1978
6. Which of the following is NOT mentioned about Hollywoodland?
   (A) It used to be the name of an area of Los Angeles.
   (B) It was formerly the name on the sign in the hills.
   (C) There were houses for sale there.
   (D) It was the most expensive area of Los Angeles.
7. The passage indicates that the sign suffered because
   (A) people damaged it
   (B) it was not fixed
   (C) the weather was bad
   (D) it was poorly constructed
8. It can be inferred from the passage that the Hollywood sign was how old when it was necessary to replace it completely?
   (A) Ten years old
   (B) Twenty-six years old
   (C) Fifty years old
   (D) Fifty-five years old
9. The word "replaced" in line 10 is closest in meaning to which of the following?
   (A) Moved to a new location
   (B) Destroyed
   (C) Found again
   (D) Exchanged for a newer one
10. According to the passage, how did celebrities help with the new sign?
    (A) They played instruments.
    (B) They raised the sign.
    (C) They helped get the money.
    (D) They took part in work parties to build the sign.

Notice that this set of questions, based on a 250-word passage, covers the comprehension of these features:

- main idea (topic)
- expressions/idioms/phrases in context
- inference (implied detail)
- grammatical features
- detail (scanning for a specifically stated detail)
- excluding facts not written (unstated details)
- supporting idea(s)
- vocabulary in context

These specifications, and the questions that exemplify them, are *not* just a string of "straight" comprehension questions that follow the thread of the passage. The questions represent a sample of the test specifications for TOEFL reading passages, which are derived from research on a variety of abilities good readers exhibit. Notice that many of them are consistent with strategies of effective reading: skimming for main idea, scanning for details, guessing word meanings from context, inferencing, using discourse markers, etc. To construct your own assessments that involve short reading passages followed by questions, you can begin with TOEFL-like specs as a basis. Your focus in your own classroom will determine which of these—and possibly other specifications—you will include in your assessment procedure, how you will frame questions, and how much weight you will give each item in scoring.

The technology of computer-based reading comprehension tests of this kind enables some additional types of items. Items such as the following are typical:

Computer-based TOEFL® reading comprehension item

- Click on the word in paragraph 1 that means "subsequent work."
- Look at the word *they* in paragraph 2. Click on the word that *they* refers to.
- The following sentence could be added to paragraph 2:

  Instead, he used the pseudonym Mrs. Silence Dogood.

  Where would it best fit into the paragraph? Click on the square ☐ to add the sentence to the paragraph.

- Click on the drawing that most closely resembles the prehistoric coelacanth. [*Four drawings are depicted on the screen.*]

Short-Answer Tasks

Multiple-choice items are difficult to construct and validate, and classroom teachers rarely have time in their busy schedules to design such a test. A popular alternative

to multiple-choice questions following reading passages is the age-old short-answer format. A reading passage is presented, and the test-taker reads questions that must be answered in a sentence or two. Questions might cover the same specifications indicated above for the TOEFL reading, but be worded in question form. For example, in a passage on the future of airline travel, the following questions might appear:

Open-ended reading comprehension questions

1. What do you think the main idea of this passage is?
2. What would you infer from the passage about the future of air travel?
3. In line 6 the word *sensation* is used. From the context, what do you think this word means?
4. What two ideas did the writer suggest for increasing airline business?
5. Why do you think the airlines have recently experienced a decline?

Do not take lightly the design of questions. It can be difficult to make sure that they reach their intended criterion. You will also need to develop consistent specifications for acceptable student responses and be prepared to take the time necessary to accomplish their evaluation. But these rather predictable disadvantages may be outweighed by the face validity of offering students a chance to construct their own answers, and by the washback effect of potential follow-up discussion.

Editing (Longer Texts)

The previous section of this chapter (on selective reading) described editing tasks, but there the discussion was limited to a list of unrelated sentences, each presented with an error to be detected by the test-taker. The same technique has been applied successfully to longer passages of 200 to 300 words. Several advantages are gained in the longer format.

First, *authenticity* is increased. The likelihood that students in English classrooms will read connected prose of a page or two is greater than the likelihood of their encountering the contrived format of unconnected sentences. Second, the task *simulates proofreading* one's own essay, where it is imperative to find and correct errors. And third, if the test is connected to a specific curriculum (such as placement into one of several writing courses), the test designer can draw up specifications for a number of grammatical and rhetorical *categories that match the content* of the courses. Content validity is thereby supported, and along with it the face validity of a task in which students are willing to invest.

Imao's (2001) test introduced one error in each numbered sentence. Test-takers followed the same procedure for marking errors as described in the previous section. Instructions to the student included a sample of the kind of connected prose that test-takers would encounter:

Contextualized grammar editing tasks (Imao, 2001)

(1) Ever since supermarkets first appeared, they have been take over the world.
A B C D

(2) Supermarkets have changed people's life styles, yet and at the same time,
A B C

changes in people's life styles have encouraged the opening of supermarkets. (3) As
D

a result this, many small stores have been forced out of business. (4) Moreover, some
A B C D B

small stores will be able to survive this unfavorable situation.
A C D

This can all be achieved in a multiple-choice format with computer scannable scoring for a rapid return of results. Moreover, not only does an overall score provide a holistic assessment, but for the placement purposes that Imao's research addressed, teachers were able to be given a diagnostic chart of each student's results within all of the specified categories of the test. For a total of 32 to 56 items in his editing test, Imao (2001, p. 185) was able to offer teachers a computer-generated breakdown of performance in the following categories:

Sentence structure
Verb tense
Noun/article features
Modal auxiliaries
Verb complements
Noun clauses
Adverb clauses
Conditionals
Logical connectors
Adjective clauses (including relative clauses)
Passives

These categories were selected for inclusion from a survey of instructors' syllabuses in writing courses and proofreading workshops. This is an excellent example of the washback effect of a relatively large-scale, standardized multiple-choice test. While one would not want to use such data as absolutely predictive of students' future

work, they can provide guidelines to a teacher on areas of potential focus as the writing course unfolds.

Scanning

Scanning is a strategy used by all readers to find relevant information in a text. Assessment of scanning is carried out by presenting test-takers with a text (prose or something in a chart or graph format) and requiring rapid identification of relevant bits of information. Possible stimuli include

- a one- to two-page news article,
- an essay,
- a chapter in a textbook,
- a technical report,
- a table or chart depicting some research findings,
- a menu, and
- an application form.

Among the variety of scanning objectives (for each of the genres named above), the test-taker must locate

- a date, name, or place in an article;
- the setting for a narrative or story;
- the principal divisions of a chapter;
- the principal research finding in a technical report;
- a result reported in a specified cell in a table;
- the cost of an item on a menu; and
- specified data needed to fill out an application.

Scoring of such scanning tasks is amenable to specificity if the initial directions are specific ("How much does the dark chocolate torte cost?"). Since one of the purposes of scanning is to *quickly* identify important elements, timing may also be calculated into a scoring procedure.

Ordering Tasks

Students always enjoy the activity of receiving little strips of paper, each with a sentence on it, and assembling them into a story, sometimes called the "strip story" technique. Variations on this can serve as an assessment of overall global understanding of a story and of the cohesive devices that signal the order of events or ideas. Alderson et al. (1995, p. 53) warn, however, against assuming that there is only one logical order. They presented these sentences for forming a little story.

Sentence-ordering task

Put the following sentences in the correct order:

A	it was called "The Last Waltz"
B	the street was in total darkness
C	because it was one he and Richard had learnt at school
D	Peter looked outside
E	he recognised the tune
F	and it seemed deserted
G	he thought he heard someone whistling

"D" was the first sentence, and test-takers were asked to order the sentences. It turned out that two orders were acceptable (DGECABF and DBFGECA), creating difficulties in assigning scores and leading the authors to discourage the use of this technique as an assessment device. But if you are willing to place this procedure in the category of informal and/or formative assessment, you might consider the technique useful. Different acceptable sentence orders become an instructive point for subsequent discussion in class, and you thereby offer washback into students' understanding of how to connect sentences and ideas in a story or essay.

Information Transfer: Reading Charts, Maps, Graphs, Diagrams

Every educated person must be able to comprehend charts, maps, graphs, calendars, diagrams, and the like. Converting such nonverbal input into comprehensible intake requires not only an understanding of the graphic and verbal conventions of the medium but also a linguistic ability to interpret that information to someone else. Reading a map implies understanding the conventions of map graphics, but it is often accompanied by telling someone where to turn, how far to go, etc. Scanning a menu requires an ability to understand the structure of most menus as well as the capacity to give an order when the time comes. Interpreting the numbers on a stock market report involves the interaction of understanding the numbers and of conveying that understanding to others.

All of these media presuppose the reader's appropriate schemata for interpreting them and often are accompanied by oral or written discourse in order to convey, clarify, question, argue, and debate, among other linguistic functions. Virtually every language curriculum, from rock-bottom beginning levels to high-advanced, utilizes this nonverbal, visual/symbolic dimension. It is therefore imperative that assessment procedures include measures of comprehension of nonverbal media.

To comprehend information in this medium (hereafter referred to simply as "graphics"), learners must be able to

- comprehend specific conventions of the various types of graphics;
- comprehend labels, headings, numbers, and symbols;
- comprehend the possible relationships among elements of the graphic; and
- make inferences that are not presented overtly.

The act of comprehending graphics includes the linguistic performance of oral or written interpretations, comments, questions, etc. This implies a process of **information transfer** from one skill to another: in this case, from reading verbal and/or nonverbal information to speaking/writing. Assessment of these abilities covers a broad spectrum of tasks. Here is a start of the many possibilities.

Tasks for assessing interpretation of graphic information

1. Read a graphic; answer simple, direct information questions. For example:
   - map: "Where is the post office?"
   - family tree: "Who is Tony's great grandmother?"
   - statistical table: "What does $p < .05$ mean?"
   - diagram of a steam engine: "Label the following parts."
2. Read a graphic; describe or elaborate on information.
   - map: "Compare the distance between San Francisco and Sacramento to the distance between San Francisco and Monterey."
   - store advertisements: "Who has the better deal on grapes, Safeway or Albertsons?"
   - menu: "What comes with the grilled salmon entrée?"
3. Read a graphic; infer/predict information.
   - stock market report: "Based on past performance, how do you think Macrotech Industries will do in the future?"
   - directions for assembling a bookshelf: "How long do you think it will take to put this thing together?"
4. Read a passage; choose the correct graphic for it.
   - article about the size of the ozone hole in the Antarctic: "Which chart represents the size of the ozone hole?"
   - passage about the history of bicycles: "Click on the drawing that shows a penny-farthing bicycle."
5. Read a passage with an accompanying graphic; interpret both.
   - article about hunger and population, with a bar graph: "Which countries have the most hungry people and why?"
   - article on number of automobiles produced and their price over a 10-year period, with a table: "What is the best generalization you can make about production and the cost of automobiles?"

6. Read a passage; create or use a graphic to illustrate.
   directions from the bank to the post office: "On the map provided, trace the route from the bank to the post office."
   article about deforestation and carbon dioxide levels: "Make a bar graph to illustrate the information in the article."
   story including members of a family: "Draw Jeff and Christina's family tree."
   description of a class schedule: "Fill in Mary's weekly class schedule."

All these tasks involve retrieving information from either written or graphic media and transferring that information to productive performance. It is sometimes too easy to simply conclude that *reading* must involve only 26 alphabetic letters, with spaces and punctuation, thus omitting a huge number of resources that we consult every day.

DESIGNING ASSESSMENT TASKS: EXTENSIVE READING

Extensive reading involves somewhat longer texts than we have been dealing with up to this point. Journal articles, technical reports, longer essays, short stories, and books fall into this category. The reason for placing such reading into a separate category is that reading of this type of discourse almost always involves a focus on meaning using mostly top-down processing, with only occasional use of a targeted bottom-up strategy. Also, because of the extent of such reading, formal assessment is unlikely to be contained within the time constraints of a typical formal testing framework, which presents a unique challenge for assessment purposes.

Another complication in assessing extensive reading is that the expected response from the reader is likely to involve as much written (or sometimes oral) performance as reading. For example, in asking test-takers to respond to an article or story, one could argue that a greater emphasis is placed on writing than on reading. This is no reason to sweep extensive reading assessment under the rug; teachers should not shrink from the assessment of this highly sophisticated skill.

Before examining a few tasks that have proved to be useful in assessing extensive reading, it is essential to note that a number of the tasks described in previous categories can apply here. Among them are

- impromptu reading plus comprehension questions,
- short-answer tasks,
- editing,
- scanning,
- ordering,
- information transfer, and
- interpretation (discussed under graphics).

In addition to those applications are tasks that are unique to extensive reading: skimming, summarizing, responding to reading, and note-taking.

Skimming Tasks

Skimming is the process of rapid coverage of reading matter to determine its gist or main idea. It is a prediction strategy used to give a reader a sense of the topic and purpose of a text, the organization of the text, the perspective or point of view of the writer, its ease or difficulty, and/or its usefulness to the reader. Of course skimming can apply to texts of less than one page, so it would be wise not to confine this type of task just to extensive texts.

Assessment of skimming strategies is usually straightforward: the test-taker skims a text and answers questions such as the following:

Skimming tasks

> What is the main idea of this text?
> What is the author's purpose in writing the text?
> What kind of writing is this [newspaper article, manual, novel, etc.]?
> What type of writing is this [expository, technical, narrative, etc.]?
> How easy or difficult do you think this text will be?
> What do you think you will learn from the text?
> How useful will the text be for your [profession, academic needs, interests]?

Responses are oral or written, depending on the context. Most assessments in the domain of skimming are informal and formative: they are grist for an imminent discussion, a more careful reading to follow, or an in-class discussion, and therefore their washback potential is good. Insofar as the subject matter and tasks are useful to a student's goals, authenticity is preserved. Scoring is less of an issue than providing appropriate feedback to students on their strategies of prediction.

Summarizing and Responding

One of the most common means of assessing extensive reading is to ask the test-taker to write a summary of the text. The task that is given to students can be very simply worded:

Directions for summarizing

> Write a summary of the text. Your summary should be about one paragraph in length (100–150 words) and should include your understanding of the main idea and supporting ideas.

Evaluating summaries is difficult: Do you give test-takers a certain number of points for targeting the main idea and its supporting ideas? Do you use a full/partial/no-credit point system? Do you give a holistic score? Imao (2001) used four criteria for the evaluation of a summary:

Criteria for assessing a summary (Imao, 2001, p. 184)

1. Expresses accurately the main idea and supporting ideas.
2. Is written in the student's own words; occasional vocabulary from the original text is acceptable.
3. Is logically organized.
4. Displays facility in the use of language to clearly express ideas in the text.

As you can readily see, a strict adherence to the criterion of assessing reading, and reading only, implies consideration of only the first factor; the other three pertain to writing performance. The first criterion is nevertheless a crucial factor; otherwise the reader-writer could pass all three of the other criteria with virtually no understanding of the text itself. Evaluation of the reading comprehension criterion will of necessity remain somewhat subjective because the teacher will need to determine degrees of fulfillment of the objective (see below for more about scoring this task).

Of further interest in assessing extensive reading is the technique of asking a student to **respond** to a text. The two tasks should not be confused with each other: summarizing requires a synopsis or overview of the text, while responding asks the reader to provide his or her own opinion on the text as a whole or on some statement or issue within it. Responding may be prompted by such directions as this:

Directions for responding to reading

In the article "Poisoning the Air We Breathe," the author suggests that a global dependence on fossil fuels will eventually make air in large cities toxic. Write an essay in which you agree or disagree with the author's thesis. Support your opinion with information from the article and from your own experience.

One criterion for a good response here is the extent to which the test-taker accurately reflects the content of the article and some of the arguments therein. Scoring is also difficult here because of the subjectivity of determining an accurate reflection of the article itself. For the reading component of this task, as well as the summary task described above, a holistic scoring system may be feasible:

Holistic scoring scale for summarizing and responding to reading

3	Demonstrates clear, unambiguous comprehension of the main and supporting ideas.
2	Demonstrates comprehension of the main idea but lacks comprehension of some supporting ideas.
1	Demonstrates only a partial comprehension of the main and supporting ideas.
0	Demonstrates no comprehension of the main and supporting ideas.

The teacher or test administrator must still determine shades of gray between the point categories, but the descriptions help to bridge the gap between an empirically determined evaluation (which is impossible) and wild, impressionistic guesses.

An attempt has been made here to underscore the *reading* component of summarizing and responding to reading, but it is crucial to consider the interactive relationship between reading and writing that is highlighted in these two tasks. As you direct students to engage in such integrative performance, it is advisable not to treat them as tasks for assessing reading alone.

Note-Taking and Outlining

Finally, a reader's comprehension of extensive texts may be assessed through an evaluation of a process of note-taking and/or outlining. Because of the difficulty of controlling the conditions and time frame for both these techniques, they rest firmly in the category of informal assessment. Their utility is in the strategic training that learners gain in retaining information through marginal notes that highlight key information or organizational outlines that put supporting ideas into a visually manageable framework. A teacher, perhaps in one-on-one conferences with students, can use student notes/outlines as indicators of the presence or absence of effective reading strategies, and thereby point the learners in positive directions.

§ § § § §

In his introduction to Alderson's (2000, p. xx) book on assessing reading, Lyle Bachman observed: "Reading, through which we can access worlds of ideas and feelings, as well as the knowledge of the ages and visions of the future, is at once the most extensively researched and the most enigmatic of the so-called language skills." It's the almost mysterious "psycholinguistic guessing game" (Goodman, 1970) of reading that poses the enigma. We still have much to learn about how people learn to read, and especially about how the brain accesses, stores, and recalls visually represented language. This chapter has illustrated a number of possibilities for assessment of reading across the continuum of skills, from basic letter/word recognition

to the retention of meaning extracted from vast quantities of linguistic symbols. I hope it will spur you to go beyond the confines of these suggestions and create your own methods of assessing reading.

EXERCISES

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

1. (C) Genres of reading are listed at the beginning of the chapter. Add other examples to each of the three categories. Among the listed examples and your additions, be specific in citing what makes certain 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. (G) Look at the list of micro- and macroskills of reading on pages 187–188. 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.
3. (C) Critique Figure 8.1 on page 190. Do you agree with the categorizations of length, focus, and process for each of the four types of reading?
4. (C) Review the four basic types of reading that were outlined at the beginning of the chapter. Offer examples of each and pay special attention to distinguishing between perceptive and selective, and between interactive and extensive.
5. (C) In Chapter 6, eight characteristics of listening were listed (page 122) that make listening "difficult." What makes *reading* difficult? Devise a similar list that could form a set of specifications to pay special attention to in assessing reading.
6. (G) Divide the four basic types of reading among groups or pairs, one type for each. Look at the sample assessment techniques provided and evaluate them according the five principles (practicality, reliability, validity [especially face and content], authenticity, and washback). Present your critique to the rest of the class.
7. (G) In the same groups as #6 above with the same type of reading, design some item types, different from the one(s) provided here, that assess the same type of reading performance.
8. (G) In the same groups as #6 above with the same type of reading, identify which of the 10 strategies for reading comprehension (pages 188–189) are essential in order to perform the assessment task. Present those findings, possibly in a tabular format, to the rest of the class.
9. (C) In the concluding paragraph of this chapter, reference was made to the "enigmatic" nature of reading as a "psycholinguistic guessing game." Why is reading enigmatic? Why is it a "guessing game"? And what does that say about the prospects of assessing reading?

FOR YOUR FURTHER READING

Alderson, J. Charles. (2000). *Assessing reading.* Cambridge: Cambridge University Press.

This volume in the Cambridge Language Assessment Series provides a comprehensive overview of the history and current state of the art of assessing reading. With an authoritative backdrop of research underlying the construct validation of techniques for the assessment of reading comprehension, a host of testing techniques are surveyed and evaluated.

Read, John. (2000). *Assessing vocabulary.* Cambridge: Cambridge University Press.

Another in the same Cambridge series, this book addresses issues in assessing vocabulary. Do not be misled by its placement in this chapter: vocabulary can be assessed through performance in *all four* skills, not just reading. A good portion of this book centers on vocabulary knowledge for reading performance, however, and therefore is recommended here. Background research and practical techniques are explored.

Nuttall, Christine. (1996). *Teaching reading skills in a foreign language.* Second Edition. Oxford: Heinemann.

This broadly based pedagogical reference book on teaching reading also offers numerous examples and commentaries on assessment of reading (which of course goes hand in hand with teaching). Most of the references to assessment deal with informal assessment, but the book also addresses formal assessment.

TESTING READING COMPREHENSION

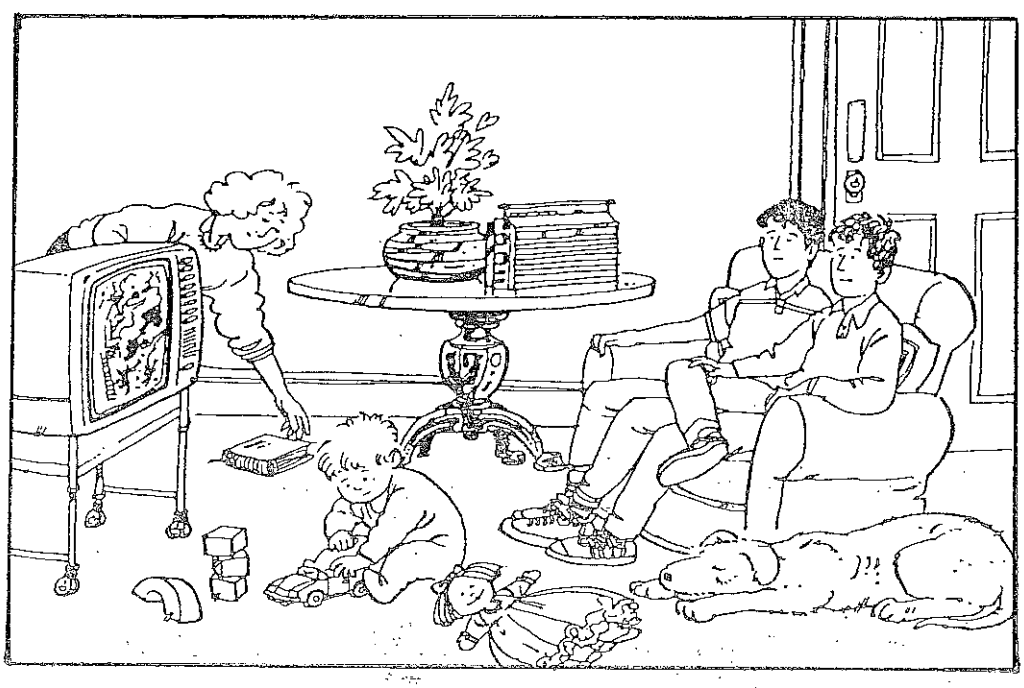
ITEM TYPES / ITEM SAMPLES FOR DISCUSSION and PRACTICE

Picture and sentence matching

TYPE 1.

Instruction :

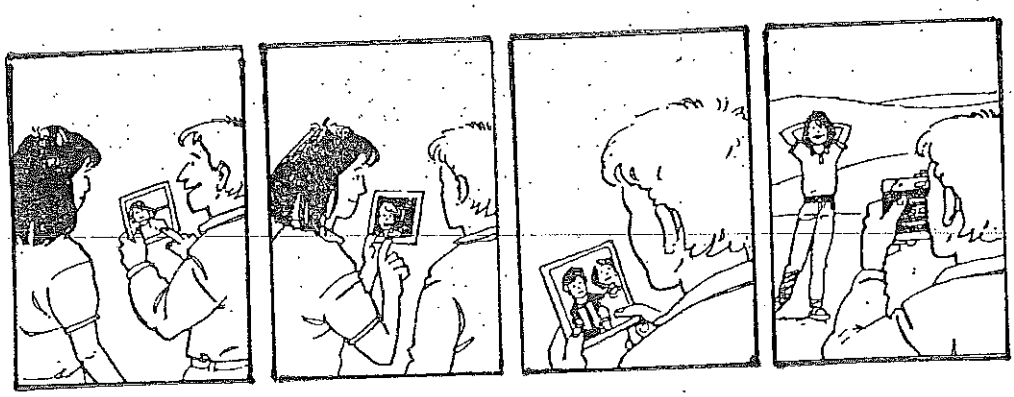
a. The dog on the floor is asleep.
b. The baby is playing with the dog.
c. The baby has just broken a toy car.
d. The television set is on fire.
e. The dog is behind the baby.
f. The woman has taken the flowers out of the bowl.
g. One of the boys is helping the woman.
h. The woman is going to pick up a book.
i. The two boys are listening to the radio.
j. The radio is on the table but the book is on the floor under the table.



TYPE 2.

Instruction:

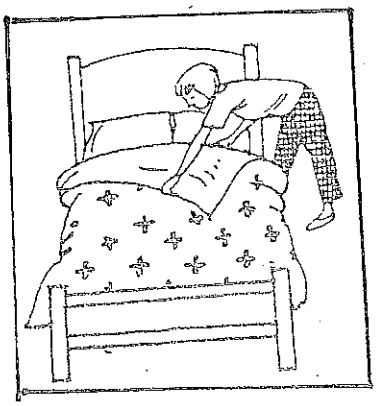
He is showing her the photograph.



TYPE 3.

Instruction :

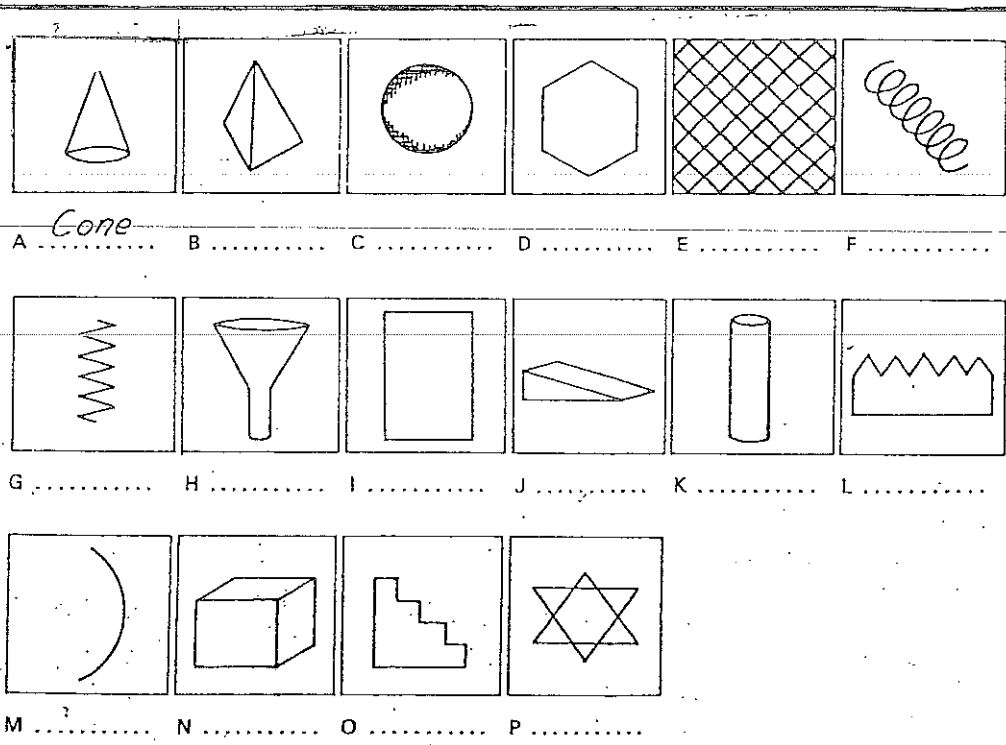
a. Jenny is throwing the ball to Peter
b. Peter is kicking the ball to Jenny.
c. Peter is throwing the ball to Jenny.
d. Jenny is kicking the ball to Peter.



a. .....
b. .....
c. .....
d. .....

MATCHING TESTS
(Intermediate and advanced stages of reading)

TYPE I : Some of the following shapes are described in the dictionary extracts. Name only those that are desribed in the extracts. The first one has been done for you.



coil /kɔɪl/ *vt,vi* wind or twist into a continuous circular or spiral shape; curl round and round: *The snake ~ed (itself) round the branch.* □ *n* [C] **1** something coiled; a single turn of something coiled: *the thick ~s of a python.* **2** length of wire wound in a spiral to conduct electric current.

cone /kəʊn/ *n* [C] **1** solid body which narrows to a point from a round, flat base. **2** something of this shape whether solid or hollow. **3** fruit of certain evergreen trees (fir, pine, cedar).

cube /kjuːb/ *n* [C] **1** solid body having six equal square sides; block of something so shaped or similarly shaped. **2** (*maths*) product of a number multiplied by itself twice: *The ~ of 5* (5^3) is 5 × 5 × 5 (*125*). □ *vt* multiply a number by itself twice: *10 ~d is 1 000.*

cu·bic /ˈkjuːbɪk/ *adj* having the shape of a cube; of a cube: *one ~ metre,* volume of a cube whose edge is one metre.

cu·bi·cal /ˈkjuːbɪkəl/ *adj* = cubic (the usual word).

cyl·in·der /ˈsɪlɪndə(r)/ *n* [C] **1** solid or hollow body shaped like a pole or log. **2** cylinder-shaped chamber (in an engine) in which gas or steam works a piston: *a six-~ engine/motor.*

cy·lin·dri·cal /sɪˈlɪndrɪkəl/ *adj* cylinder-shaped.

el·lipse /ɪˈlɪps/ *n* [C] regular oval.

el·lip·tic /ɪˈlɪptɪk/, **el·lip·ti·cal** /-kəl/ *adj* shaped like an ellipse.

funnel /ˈfʌnəl/ *n* [C] **1** tube or pipe wide at the top and narrowing at the bottom, for pouring liquids or powders through small openings. **2** outlet for smoke of a steamer, railway engine, etc. □ *vt, vi* (-ll-, *US* -l-) (cause to) move (as if) through a funnel.

lat·tice /ˈlætɪs/ *n* [C] framework of crossed laths or metal strips as a screen, fence or door, or for climbing plants to grow over: *a ~ window.* **lat·ticed** *adj*

pyra·mid /ˈpɪrəmɪd/ *n* [C] **1** structure with a triangular or square base and sloping sides meeting at a point, esp one of those built of stone in ancient Egypt. **2** pile of objects in the shape of a pyramid.

sphere /sfɪə(r)/ *n* [C] **1** form of a globe; star; planet. ***music of the spheres,*** (*myth*) music produced by the movement of heavenly bodies which men cannot hear. **2** globe representing the earth or the night sky.

spheri·cal /ˈsferɪkəl/ *adj* shaped like a sphere.

wedge /wedʒ/ *n* [C] **1** V-shaped piece of wood or metal, used to split wood or rock (by being hammered), to widen an opening or to keep two things separate, ***the thin end of the wedge,*** (*fig*) a small change or demand likely to lead to big changes or demands. **2** something shaped like or used like a wedge: *~ heels (on shoes).* □ *vt* fix tightly (as) with a wedge: *~ a door open,* by placing a wedge under it. *be tightly ~d between two fat women on the bus.*

TYPE II : Matching tests based on a text and the dictionary.

Settlements

Enclosed hut groups are characteristic settlements in the area and remains of more than a hundred still exist. The open settlements, the villages of predominantly unenclosed huts, are not numerous but only a dozen groups have sufficient numbers of co-ordinated huts to be described as villages. Though there may be some walls in these villages, they are only fragmentary.

Many enclosed settlements have disappeared but one still extant is Rides Rys. It consists of two enclosures, one roughly square and a larger one roughly oblong in plan with a shared wall. An area of six acres was enclosed containing more than thirty buildings. The large enclosure, as in the case of other multiple settlements, had been built on to the smaller and indicates an expanding community.

The following dictionary definitions are four words used in the passage Settlements. Write the words from the passage next to the appropriate definition.

1. *having or involving more than one part as an individual* .....
2. *serving as repersentative* .....
3. *becoming greater in size* .....
4. *is a sign of* .....
5. *in most cases* .....
6. *very near to; approximate* .....

TYPE III : The matching of information

It is important for each student to obtain at regular intervals a rough idea of his or her progress. How are goals being achieved week by week? Methods of continuous assessment of students' work are replacing examinations – or parts of examinations – on certain courses. There are still doubts about the advantages of continuous assessment in the learning process but, if applied with care and discretion, continuous assessment can be a far more valuable means of assessing standards than an examination. Provided that methods of continuous assessment do not impart a feeling of tension and strain, they can be used to guide students in their work and to inform them of the progress they are making. If no means of continuous assessment is available, students should attempt to evaluate and summarise their progress very briefly week by week. Clearly, such an attempt is more difficult in a subject which teaches skills (e.g. learning a language, playing a musical instrument) than in a content subject (e.g. history, chemistry). Even as far as skills are concerned, however, it is a simple matter for students to go back to an old exercise and do it again. The ease with which they can do what previously seemed a difficult exercise is often quite remarkable.

1. Instead of talking about using such methods carefully and wisely, the writer talks about applying them .....
2. Instead of saying that it is harder for the students to assess progress made in learning a language than in history, the writer says that .....
3. Instead of referring to ways of measuring students' progress at regular intervals, the writer uses the phrase .....
4. Instead of saying that students should try to assess and report briefly on their progress every week, the writer says that .....
5. Instead of saying that continuous assessment is useful unless it makes students feel upset and worried, the writer says that it is useful .....

FALSE/TRUE READING TESTS

TYPE 1. Those which are independent of a reading text.

Instruction:

1. When ice melts, it turns into water. T F
2. Unlike tigers, deer can run fast. T F
3. Married couples get divorced when they get along well with each other. T F
4. All people dream about being famous T F
5. People eat a lot when they feel depressed. T F
6. Spagetti is more delicious than rice. T F
7. Green is a colour which makes people relaxed. T F
8. Rice contains more carbonhydrate than sugar. T F
9. ..... T F
10. ..... T F

TYPE 2. Those which depend on a text (*)

MULTIPLE-CHOICE ITEMS: Short texts

Type 1.

- The eyes are wonderful teachers – even musicians, who deal with sound, learn as much by { *doing, playing, watching, practising* } as by listening.
- The housewife who could not afford to buy clothes woiuld spend hours at her spinning wheel, spinning her wool into yarn – a job which took little skill but required a lot of { *ability, patience, talent, wisdom* } and was done by the fireside during the long winter evenings.

Type 3. One item based on a short paragraph.

- 1. The president was talking to a young woman in the crowd when Tim suddenly caught sight of a man standing several yards behind her. The man had something in his hand: It was a short stick.

What made Tim notice the man in the crowd ?

A. He was very close to Tim.
B. The president was talking to him.
C. He was standing in front of the woman.
D. He was carrying a stick.

- 2. There were only two ways of moving along the narrow ledge: face outwards or face to the wall. I concluded that even the smallest of bottoms would push a person with his back to the wall far enough out to overbalance him and so, with arms outstreched in the direction I was heading. I inched my day along.

The writer managed to cross the narrow ledge by...

A. crawling along on his knees with his arms stretched out in front of him.
B. moving sideways inch by inch with his back to the wall
C. working his way forward on his stomach with his face almost touching the ledge.
D. walking slowly with his face and stomach close to the wall.

Type 4. Reference- word items

(1)The grizzly bear roams some 12 million acres in rugged parts of the United States. And this great bear still roams our
(2)imagination at will: It is part of its natural habit.

The word **it** in line 3 refers to ..

A. "the United States" (line 2)
B. "this great bear" (line 2)
C. "our imagination" (line 3)
D. "its natural habit" (line3)

Type 5. Items about numerical information

TASK. Read the following paragraph and construct a multiple-choice item based on the numerical information given in the passage.

Latest reports from the northeast provinces state that at least sixteen people losr their lives in Saturday's floods. A further nine people, mostly children, are reported missing, believed dead. Seven small boys, however, had a miraculous escape when thet were swept onto the branches of small tall trees.

.....
.....

A. **B.** **C.** **D.**

TASK. Discuss the following items which are supposed to be based on reading passages.

- Memorizing is easier when the material to be learnt is ...

A. in a foreign language
B. already partly known
C. unfamiliar and not too easy
D. of no special interest

- According to the passage, what do some people think there should be outside a modern city ?

A. Buses
B. Car Parks
C. Office buildings
D. Taxis
E. ALL OF THESE

- Paul did not expect to see Sue because ....
  A. he did not know she was at the party.
  B. no one knew she had left the district
  C. he hadn't seen Jane teaching her to drive
  D. he didn't realize she was ill.

- The curriculum at the new college is a good one in many ways because it ...

  A. includes many science programs
  B. offers a well-balanced programme in both humanities and the sciences.
  C. is realistic.
  D. Consists of useful techical subjects.

- We may infer from the paragraph that people...

  A. all need the same kind of rest.
  B. do not ususally need rest.
  C. rest much more than they should
  D. do not all rest in the same way.