

## TESTING LISTENING COMPREHENSION IN JAPANESE UNIVERSITY ENTRANCE EXAMINATIONS

Gary Buck

### *Abstract*

Several major Japanese universities have recently included or are shortly to include an English listening comprehension section in their entrance examinations. It may be expected that a large number of other colleges will follow suit. This paper examines the nature of listening comprehension, and the problems of testing it, with emphasis on those problems relevant to constructing university entrance examinations. Section 1 looks at the importance of college entrance examinations in Japan, and their effect upon classroom instruction. Section 2 examines the nature of listening comprehension; reviews theories and research, and provides criteria for creating and evaluating listening texts. In section 3, the most common methods of testing listening comprehension are discussed in terms of these criteria, and in terms of practical implementation and washback effect in the classroom. Practical recommendations are made for constructing comprehension tests suitable for Japanese university entrance examinations.

Recently a number of universities, including Tokyo University, have added an English listening test to their other entrance examinations. Many other colleges are likely to follow this lead and include listening tests in their own entrance examinations. However, listening comprehension is an extremely complex process, about which little is known, and thus the testing of listening comprehension presents considerable problems. The purpose of this paper is twofold. First, to examine some of the complex issues involved in understanding listening comprehension, how these relate to the construction of listening comprehension tests, and to offer some practical suggestions about the advantages and disadvantages of most of the common methods of testing it. Of course, there are no simple, instant solutions to such

Gary Buck is an assistant professor at Osaka Meijo Women's Junior College, and is currently working towards a doctorate in testing second language listening comprehension at Lancaster University.

complex problems. Rather, it is necessary to examine the issues and make the best decisions possible in the light of the available evidence. The second purpose of this paper is to try to stimulate and encourage practical and helpful discussion of these issues, among both language teachers and testers.

### 1. The English Entrance Examinations

Japan is a country in which the entrance examination reigns supreme. It is almost impossible to overstate the influence of these examinations on both the educational system as a whole, and the day-to-day content of classroom teaching. Their importance in the lives of young people is such that almost all future social and economic advancement is dependent on the results of these entrance examinations.

#### *1.1 Constraints on Entrance Examinations*

However, despite the immense importance of these examinations, there are a number of practical constraints on their production which considerably influence the nature of the tests produced. With the exception of the first part of the entrance examination to national universities, entrance examinations are usually made by each individual college, and colleges generally have two or three different examinations every year, which leaves very little time for development. The actual examination is usually made by a committee of teachers in each college, very often on an annually rotating basis. Due to the large number of candidates and the necessity (or custom) of completing all the marking in one long session the same or the next day, the tests must be quickly and easily scorable by a large group of teachers who may not have been involved in the test-making process. Security is a continual problem, considering the importance of the results, and it is generally felt to be impossible to pre-test items in order to ascertain whether they function as the item writer intended. Nor is it possible to use items which seemed to work well in past years, as most schools feel they must publish each year's entrance examination for the benefit of future applicants. A further important constraint, which should not be overlooked, is the feeling that the examination reflects on the prestige of the college making it. This means in practice that a test should look difficult, to give the impression that the college has a very high standard.<sup>1</sup>

### *1.2 The Nature of the Tests*

Given these considerable constraints, the actual tests themselves have to be put together quickly by a committee of teachers with little specialized training in Educational Measurement as a separate academic discipline. In such a situation most teachers see little alternative but to try either to produce a different version of the tests made in previous years, or imitate the tests of more prestigious institutions.

There has developed something which may be easily recognized as a typical Japanese college entrance examination, which seems to be based on no clearly stated view of language or language acquisition. Furthermore, there is rarely any attempt to test language in use, and the items used are often of a type not currently recommended by language testing researchers (Heaton, 1975, p. 186; Buck, in press). Finally, these tests are seldom subjected to any standard statistical analyses or other validation procedures; it is generally assumed by both test makers and test users that tests made by experienced teachers are automatically good tests. According to Spolsky's classification, these tests are clearly in the "pre-scientific stage" of language testing (Spolsky 1975).

### *1.3 Effects on Classroom Teaching*

There is a natural tendency for both teachers and students to tailor their classroom activities to the demands of the test, especially when the test is very important to the future of the students, and pass rates are used as a measure of teacher success. This influence of the test on the classroom (referred to as *washback* by language testers) is, of course, very important; this washback effect can be either beneficial or harmful.

Most educators would probably agree that the content of classroom instruction should be decided on the basis of clearly understood educational goals, and examinations should try to ascertain whether these goals have been achieved. When the examination does that, it forces students and teachers to concentrate on these goals, and the washback effect on the classroom is very beneficial. However, if the examinations are not testing these goals, students, who usually have their sights set on the examination, may pass the tests but totally fail to attain the basic goals set by educational planners. In such a case the washback effect is very negative indeed. It is perhaps here that Japanese entrance examinations are most destructive.

Many English teachers in Japan want their students to be able to use English as a means of communication. However, as long as entrance

examinations consist mainly of such items as discrete-point grammar questions and translation, then students are obviously going to concentrate their efforts on learning to answer grammar questions and do translations, despite the wealth of evidence which suggests that such activities will not lead to successful language learning.<sup>2</sup> (Krashen, 1982; Savignon, 1983; Richards & Rogers, 1986). If we want our students to learn to communicate in English, then we must give them tests which require them to process communicative English.

There are probably many reasons why most Japanese high school graduates cannot use English for even the most basic purposes, despite receiving hundreds of hours of classroom instruction, but surely one of the most important is the washback effect of entrance examinations on the classroom.

### *1.4 Possibilities for the Future*

The new trend towards including a listening comprehension section in entrance examinations seems to offer an opportunity to make examinations which test "real" English and will thus have a positive washback effect on the classroom. Ideally the new listening tests should fulfill two roles. First, they should provide an accurate and fair measure of the English listening ability of testees. Second, they should have a positive washback effect on the classroom, so that when students study for the test, they will automatically have to engage in activities which will lead to effective language acquisition. Although listening comprehension is an extremely complex process, it is the writer's opinion that both these aims can be attained with a fair degree of success if test makers are prepared to confront the considerable problems involved.

It is obviously impossible to measure anything without first establishing exactly what is being measured. Therefore, in order to come to some sort of understanding of how listening tests should be made, the most obvious necessity is first to examine the nature of listening comprehension itself. This will provide a theoretical basis for teachers and test makers to develop their own critical standards, and thus enable them to evaluate listening tests for themselves.

## **2. Listening Comprehension**

Unfortunately, it is very difficult to come up with a simple definition of what listening is, and what listening comprehension means. Instead

we have to consider research from various fields and try somehow to bring this together and arrive at some workable conclusion.

A number of different, but related, skills are involved in listening. The listener must take in a stream of sound and somehow convert this into meaning. To do that it is clearly necessary to have knowledge of the language. This knowledge obviously includes the phonological, morphemic, syntactic and semantic rules of the language, as well as cohesive devices, text types, etc., and the ability, in practice, to apply such knowledge rapidly and automatically. Researchers also emphasize the importance of pragmatic and general world knowledge in language comprehension. Naturally in such a complex area different theorists or researchers produce different classifications of these sub-skills. The key area of disagreement, however, concerns the relationship between these skills, and how they affect each other. There are two main lines of thought: 1) that language processing is only bottom-up, or 2) that it is both *top-down* and *bottom-up*.<sup>1</sup>

### *2.1 Bottom-up Processing*

The bottom-up approach sees language comprehension as a process of passing through a number of consecutive stages, and in each stage one particular sub-skill is utilized, starting with the lowest level and moving up to higher levels of processing. First, at the lowest level, the acoustic input is decoded into phonemes, and then the information obtained is used to identify individual words, after which processing continues on to the next stage, the syntactic level. Only after the syntactic level is completed is the semantic content of the utterance extracted. Processing is thus seen as occurring on a number of different levels, starting from the bottom and working upwards through one level to the next, in serial order. The output of each level thus becomes the input for the next higher level. Of course, this is an oversimplification of a number of different, extremely complex, psycholinguistic theories (for a review, see Dirven & Oakeshott-Taylor, 1984, 1985). However, the important thing to note about the bottom-up approach to language comprehension is that the results of processing on a higher level are not available for use at lower levels. It is, as it were, a one-way street.

## 2.2 Top-down Processing

Advocates of the top-down approach to language processing generally accept the importance of lower level information, and also the fact that processing takes place on a number of different levels. However, they suggest that processing does not occur in a fixed order, from bottom to top, but rather, different types of processing may occur simultaneously, or higher level processing may take place before lower level processing. Thus, the results of higher levels of processing may be available to facilitate processing at lower levels. This means that they think it is quite possible in practice to understand the meaning of a word before decoding its sound. In almost all situations, linguistic and non-linguistic, our experience leads us to have expectations about what will happen next. These expectations, or hypotheses, may be very precise or rather general but nevertheless we almost always have some idea about what is likely to come next when we are listening. In such cases it is not necessary to utilize all the lower level information available to us, we can just take in enough to confirm or reject our hypotheses. Comprehension, then, becomes largely a process of hypothesis generation and testing. For example, if we hear the following uncompleted sentence:

[1] She picked up the gun, aimed and \_\_\_\_\_ (Grosjean, 1980, p. 281)

we probably need very little acoustic information to understand that the final word will be *fired*. When we listen, we will hypothesize that the last word will be *fired* and then probably process only enough of the acoustic input to confirm our hypothesis, or, if we feel really sure, we may not even bother to listen to the last word at all. Similarly, when we part from a friend, we may hear a word of farewell, not so much by processing what he says, but because he is waving to us and saying something as he walks away.

Indeed, research has shown that our expectations are so strong that they can often lead us to "hear" things which were quite different from what was actually said. One cognitive psychologist, on the basis of his research results, concluded that, "contextual information is able to control the lexical interpretation assigned to a given acoustic-phonetic sensory analysis" (Marslen-Wilson, 1980, p. 49). Which means that what we hear is often a product of what we expect to hear. Bruce (1958), in a famous experiment, found that when subjects were asked to repeat a sentence heard against a background of noise, what they repeated was

greatly influenced by what they had been informed was the topic of the sentence. When told that the sentence would deal with 'sport', the listener correctly repeated *I tell you that our team will win the cup next year*. However, when informed that the topic was 'health', the listener heard the same sentence as *I tell you that our team has been free from injury all this year*, and when told that the topic was 'weather' the listener heard *I tell you that I see the wind in the south next year*.

Over the last few years the evidence for the top-down nature of language processing has become stronger and stronger, such that now in applied linguistics it is normal to discuss comprehension in terms of top-down processing; that is, as an interactive process which simultaneously utilizes information from a number of different sources. Faerch and Kasper (1986, p. 264) suggest that comprehension is a process relying on three types of information: 1) linguistic and other communicative input; 2) the listener's linguistic and general world knowledge; and 3) information from the context and the conversation up to that point. This position has been facilitated to a great extent by research into reading. Reading and listening obviously share much in common, in that they are both different aspects of the process of language comprehension. While they do clearly differ in a number of crucial respects, it does seem reasonable for those interested in listening to look at research into reading, and consider how much of this is applicable to listening.

### *2.3 Schema Theory*

There has been a considerable amount of research into reading and it would be impossible to review it all here. However, the trend over the last few years has been to see reading less and less as a data-driven process, and more and more as a hypothesis-driven process (Alderson & Urquhart, 1984; Smith, 1985; Devine, Carrell, & Eskey, 1987). We use our knowledge of language and the world to generate hypotheses about what we will read and then extract enough data from the printed page to confirm or reject these hypotheses. The reason we can read about something we are familiar with much faster than something which is very new to us is because our knowledge helps us to guess what is contained in the text, and we thus need to take far less from the text itself. We provide more of the information ourselves. The recognition of this fact led Goodman (1970) to describe reading as a "psycholinguistic

guessing game." For each word in the language, or situation in which we find ourselves, there are a whole host of other things which are associated with it in our minds. These associations, or *schema*, form the basis of many of our expectations of what we will read (Rumelhart & Ortony, 1977; Rumelhart, 1980; Carrell, 1987). Not only that, but common situations in our lives tend to follow a reasonably predictable pattern. A visit to a restaurant usually involves things such as looking at the menu, ordering the meal, eating it and then paying for it. This pattern forms a sort of universal script for a restaurant "story" (Schank & Abelson, 1977).

These schema (scripts can be regarded as complex schema) are one of our ways of ordering our experience of the world. They seem to play a crucial role in the process of hypothesis generation which is so central to language comprehension. The context thus becomes an important part of the listening comprehension process, because it provides important clues to help us understand the meaning. Obviously, if we fail to provide a context, we rob students of important resources they would normally use in the listening comprehension process. In normal daily language use, we virtually never have to decode utterances by using only bottom-up information, devoid of the context which enables us to use top-down processing strategies. Therefore, if we ask students to decode short decontextualized sentences, we are not testing listening comprehension at all, but asking students to engage in a very unnatural activity which seems to be confined largely to the second language classroom.

There is as yet very little research into the effects of schema on listening, but there is every reason to suggest that they play just as important a part in listening as they do in reading. Unpublished research by the author has indicated that schema are tremendously important in listening. A number of students were asked to listen to a selection of different passages. While listening the tape was stopped at a number of points and the students were asked to retrospect on their listening processes. Results indicated that students found it much easier to comprehend passages which accorded well with their own knowledge, even when it seemed clear that the language used was well within their linguistic capabilities. All listeners reported creating images in their minds, which were far more detailed and complex than the descriptions contained in the texts they were listening to. When asked where these images came from, they indicated things they had learned at school, films they had seen, or books they had read. In other words, they



came from their schema associated with the descriptions they heard. Listeners often had strong expectations about what would come next in the passage they were listening to, so strong in some cases that they prevented the listener from hearing what was actually said when it differed from their expectations.

The analysis of this data is still in progress, but one thing is already very clear. Although lower level acoustic and phonetic input is, of course, tremendously important to speech processing, comprehension of language in context utilizes a considerable amount of top-down information.

There are many examples in the data of top-down processing failing for some reason, resulting in a lack of comprehension, even though there appeared to be quite enough linguistic input for comprehension to take place. One striking case shows the influence of expectations. A student was listening to a story about a girl in Africa who asked a witch-doctor to solve a number of burglaries. The student had lived in the United States and was expected to understand the story with ease. However, she didn't know the word *witch-doctor*, so she made the sensible assumption that this was some sort of detective. Based on this wrong assumption, she naturally expected the witch-doctor to proceed like any normal detective. However, instead, he performed a number of magic actions to find out who the criminal was. The description of these events should have presented no difficulty to a student of her English ability, yet she totally failed to "hear" clues that should have made it quite clear to her that this was not a normal investigation, but a series of magic rituals. Based on her expectations, she "heard" her own story of a criminal investigation, quite different from the one she listened to, and remained convinced to the end that she had understood the passage quite well.

### 2.4 Pragmatics

There are other reasons why lack of context often inhibits the comprehension of language. This is because even if we have understood the basic semantic meaning of an utterance, we often need to know the context in order to understand what was meant. Leech (1983, p. 4) illustrates this by the different uses of the verb to *mean* in the following sentences:

[2] What does X mean?

[3] What do you mean by X?

The real meaning of what is said, the pragmatic meaning, is often very different from the surface meaning, especially in interactive conversation. To give a simple and often used example, when someone says:

[4] Can you pass the salt?

they are not usually asking about our ability to pass the salt at all, but are making request that we do so. Similarly,

[5] Do that again and I'll thump you.

has the appearance of a command to do something again, but as we all know, in actual fact it is an injunction not to do it again, or maybe it's even an invitation to a fight. Only the context will tell us which.

[6] I'd keep my mouth shut if I were you.

could vary in meaning along a continuum from being a piece of friendly advice all the way to being a murder threat. Of course, the intonation would probably help us understand to some extent, but to understand the pragmatic force of this fully we would need to know something about the participants, the relationship between them, the nature of the information the speaker is referring to, and the conversation that preceded this. The context in which an utterance is used not only provides the necessary conditions for engaging in top-down processing, but also provides a background against which the semantic value of the utterance can be interpreted.

### *2.5 Reasonable Interpretation*

If the spoken message doesn't accord with our knowledge of the world, then we are going to have serious problems understanding and interpreting what we hear. Listeners' comprehension will vary depending on how well the spoken message they are listening to does, or does not, accord with their own knowledge, and hence, how much it agrees with or differs from their expectations. A talk on a subject about which we know nothing, all other things being equal, will be more difficult to understand, and further, a text which in some way violates or contradicts our expectations will be even more difficult to understand and could cause considerable confusion even though the language may not be particularly difficult.

Where high levels of new information in a listening passage make hypothesis generation more difficult, listeners will tend to be selective,

resulting in different listeners extracting different things from the same text depending on how well the new information fits in with, or supplements, their own current state of knowledge. Also, different listeners often have different motives for listening, have different interests and different needs. Listeners will pay more attention to these features of a text which they think are more interesting or more relevant to their view of what is important to them in the passage. This is a very important point because "different listeners will reasonably extract different parts of the text as more 'salient' to them. . . , and so build their mental representations of 'what the text was about' around rather different structures" (Brown & Yule, 1983, p. 100).

The corollary of all this is that there is no such thing as one "correct interpretation" of a spoken text. Of course, in practice there are likely to be some interpretations which we would all consider so wide of the mark that we could call them wrong (for example with the witch-doctor story above), and some interpretations which are generally agreed to be better than others. But it must be understood that each listener has to construct a personal interpretation of what we heard, and what is comprehended or understood may well differ from listener to listener. It is therefore more appropriate to talk of "reasonable interpretations" rather than "correct interpretations."

### 3. Testing Listening Comprehension

Given the complexity of the listening process, it is easy to see that many tests which are called tests of listening comprehension are not really testing listening comprehension in the full sense at all. Indeed, Brown and Yule express the problem thus:

. . . we find existing approaches to the assessment of listening comprehension based on a very insecure theoretical notion of what "comprehension" means. It is by no means clear that a great deal of what is currently tested in listening comprehension tests is necessary, or relevant, to the process of understanding the communicative event which the student has listened to. (1983, p. 100)

It is indeed true that there is, as yet, no such thing as the ideal listening test. There are, however, a number of different types of listening test, each with its own advantages and disadvantages. The type of test preferred will depend largely on how individual test makers evaluate these advantages and disadvantages in the light of their own requirements, which is one reason why test users should try to develop a

good theoretical understanding of listening comprehension. However, whichever testing method is chosen, it will be necessary to make compromises. For convenience, these different listening tests can be divided into two main approaches. They are the *process approach* and the *product approach*.

### 3.1 The Process Approach

The first of these, the process approach, attempts to identify the various sub-skills or processes used in listening and then tries to assess whether the testee has mastered these or not. This is not so easy, as we have already seen that the process of listening is complex and can vary according to the world knowledge and interests of the listener. However, we do know that both listeners and readers use their linguistic and world knowledge to generate hypotheses about the language they are comprehending. They do not need to get all the information (either graphic or acoustic) contained in a message in order to comprehend, because the language contains more information than is necessary. It is, in other words, highly *redundant*, and we can use that fact to test language comprehension. We can give the students a passage in which some of the information is missing (words or parts of words) and reasonably expect them to understand it. And if they understand it, we can also reasonably expect them to be able to replace the missing information. There are, in fact, tests which use this approach, they are generally referred to as *tests of reduced redundancy* (Spolsky, 1971; Gradman & Spolsky, 1975), and have found favour with a number of testing researchers (Cohen, 1977; Oller, 1979). This, of course, is the theoretical basis of the widely used cloze test.

#### 3.1.1 Noise Tests

There are a number of ways of creating reduced redundancy listening tests. One of these is the noise test, in which students hear a passage which has been mutilated to some extent by the addition of background noise to the recording. This is usually "white" noise, that is noise which covers most of the frequency range of the spectrum, a sort of continuous hiss. Of course, the text is partly masked by the noise, thus reducing the redundancy of the language, but it is expected that testees will be able to use their linguistic knowledge to recreate the original text (Spolsky, 1971; Johansson, 1973; Gradman & Spolsky, 1975; Cohen, 1977). There

are two main ways of recording students' responses. In the first, students are asked to record what they hear by repeating it into another microphone while they are listening to the original recording. This obviously presents practical problems for mass-testing situations such as Japanese entrance examinations. The other method is to insert pauses into the recording during which students write down what they have heard. This method would be quite practical for entrance examinations in Japan, but would obviously require students to have reasonable writing ability. We may find that some students who received low scores did so because of deficiencies in their writing ability rather than in their listening ability. In other words, the listening test scores could become contaminated by writing ability.

It seems to the writer that there are some very serious objections to noise tests. The first of these concerns their theoretical basis. The basic idea is that language is redundant and so students are expected to use their linguistic knowledge to restore the original text. However, we have established that listeners use far more than just their linguistic knowledge to comprehend, and yet the noise test seems to provide little opportunity to use nonlinguistic sources of knowledge during comprehension. Another serious objection concerns the washback effect of these tests on the classroom. Students studying for such tests will probably feel they have to spend valuable class time practising listening to recordings mutilated by random noise. This would be disastrous, as our aim is to have students learn to listen to English as it is really used for communicative purposes by English speakers. A third practical objection is that if we are going to ask students to write down what they hear, as we probably must, we are more or less giving them a dictation, in which case we might as well give them a normal dictation, which probably works just as well, is easier to make, and is far more acceptable to most teachers and students.

### *3.1.2 Listening Cloze Tests*

Some test makers have used listening cloze tests, in which the students usually get a written passage from which certain words have been replaced by blanks, as in the standard written cloze test, and then they listen to a recording of the passage and try to fill in the blanks from what they have heard. There are a number of problems with this technique. One is that students can often treat the passage as a normal

cloze test, and fill in the blanks even if they didn't hear the passage very well, in which case it is no longer a listening test at all. One way around this is to put the blanks on high-information content words, which tend to be more difficult to guess from linguistic knowledge (Henning, Gary, & Gary, 1983). But even then it doesn't make a very satisfactory listening test, as students usually find themselves simply listening for individual words rather than trying to understand and interpret a passage. In many cases such a test would be little different from a word recognition test. In terms of washback on the classroom the objections to the noise test seem to be equally applicable to the listening cloze.

There does seem, though, to be one form of listening cloze test which involves more natural listening activities and is thus likely to have a beneficial washback effect. That is to make a cloze passage on a summary of the original text. It may be argued, of course, that this is no longer a test of reduced redundancy, but that doesn't seem very important if the method produces good results. The testing procedure is quite simple. Students are given a summary of the passage they are going to hear, in which some of the important content words have been replaced by blanks. After looking at the passage for a while, so they have some idea what they are listening for, they listen to the passage. They should be discouraged from writing while listening. Their task is to use their understanding of the passage to fill in the blanks. One mark is awarded for each blank completed with an acceptable word.

Obviously, this technique would only work well with those texts which are amenable to summarizing. It seems particularly suitable for narrative texts, although it could probably be used with other text types. It avoids the problem of students just listening for single words, and if done well, could require them to use higher level processing skills, such as making inferences about the main point of the passage, the relationship between different events in the story, or the overall significance of certain parts of the passage. As the blanks are put on words which have a high information content, it is highly unlikely that students would be able to guess them without understanding the passage.

The thing that requires most care with this testing technique is making good summaries. It is not quite as easy as it seems. To those who write the summary it often seems obvious what should go into each blank, but to others this is not always the case. It is vital that summaries be checked by other teachers or experts to ensure there is no ambiguity

and it is always clear what information is required in each blank (Alastair Pollitt, personal communication). Although there is little research evidence on this technique, it would seem to be one which has much to recommend it, and further research is obviously called for.

The writer has used this technique in Japan with a number of English listening tests in which the summary was written in Japanese. First, a suitable narrative text was selected. Then a number of native speakers (Japanese teachers of English) were asked to write a summary of the story in Japanese.<sup>3</sup> It was found that these differed a little from each other, and so a composite of these summaries was produced in order to try and get the advantages of each. Then references to those events in the narrative which were considered central to the whole story were replaced with blanks. Because of the doubt in Japanese about what exactly constitutes a word, it was decided that deletions would not be restricted to one word, instead short phrases were deleted (instructions were given on the test which made this very clear). These passages with the blanks were then given to a number of students who were asked to try to fill in the blanks without hearing the actual story, in order to check whether the blanks could be filled in from general or linguistic knowledge. It was surprising that, despite the greatest care, some blanks could indeed be guessed by a number of students without ever listening to the story, and so the summaries were remade without those blanks, and the test was finished.

The results were encouraging. Analysis showed that such tests had acceptable reliability as estimated by Cronbach's Alpha, generally comfortably over .80.<sup>4</sup> When one of these tests was included in a battery of comprehension tests administered to over 400 Japanese college students, it correlated closer with other tests of listening comprehension than it did with a test of reading comprehension which used the same testing method. This is an important result, because it shows that this test was functioning as a listening test, rather than just as a fill-in-the-blank test. These tests, thus, seem to have had acceptable reliability and validity.

The use of clozed summaries of listening passages as a testing method seems to have much to recommend it. They apparently have a reasonably firm theoretical base, practical trials have produced encouraging results, and they look sufficiently serious and academic for inclusion on college entrance examinations. Furthermore, the washback effect on the classroom is likely to be quite beneficial if

students realize that they have to understand passages in their entirety to complete the tests.

### 3.1.3 Dictation

One test type which is in many ways related to the reduced redundancy tests and which has been very widely used is the dictation. This was criticized by Lado (1961, p. 34) as being little more than a test of spelling, but work in the 1970s has indicated that dictation can be a good measure of general language proficiency (Oller, 1979; Oakeshott-Taylor, 1977). Basically, the general idea is that although a passage may be redundant for first language users, it will be much less so for second language users. Thus, the extent to which second language users can utilize the redundancy in the language is a good measure of their language ability. For those who don't know the technique, dictation usually involves hearing a passage twice. The first time the passage is played straight through, and students just listen and try to understand. The second time they hear it, the passage is broken into a number of short sections with a pause between each section. During that pause students have to write down what they have just heard.

Simply listening to a passage of English and then writing it down requires the listener to engage in many of the activities we normally use in speech perception. Of course, the length of each section is very important. It requires very little language ability to repeat or write down a single word heard in isolation. However, if the length of the sections is increased, the students are required to rely less and less on their short-term memory, and more and more on their language processing ability. This is because there is a limit to the capacity of short-term memory, which seems to be restricted to about seven units of information (Miller, 1956). These units are often referred to as *chunks* and the size of these chunks is not fixed. This means that most students will be able to keep up to about seven words in short-term memory, assuming they can decode them in the first place. But more advanced students will be able to use their language ability to "chunk" these words into meaningful units, such as phrases or idea units, and then retain about seven of these units in memory. Those who have enough language ability to chunk the language can thus reproduce much longer sections of the text than those who don't. In the light of this, it is important, when constructing dictation tests, to ensure that each of the sections of text is long enough



to place a certain amount of load on the short-term memory of the students. This will then force them to use their linguistic knowledge to chunk the material.

The size of each section of text between pauses can vary, but probably the best way is to make them of differing lengths, varying from about four or five words up to about 12 words or so. The ideal length will depend on the ability of the students and the nature of the text used. Experience is the only real yardstick. Ideally, the sections near the beginning of the test should be shorter, and they should get progressively longer through the test.

There are a number of ways of scoring dictations. But it is important to remember that they are not designed to be tests of spelling, and so spelling mistakes should be ignored in cases when it is obvious that the mistake is indeed a simple spelling mistake (not always an easy decision). It is common to give one mark for each correct word. This can be done by adding together all the correct words in each section. Words out of order or omitted are marked as wrong. This is an easy way to mark, but unfortunately it doesn't take account of intrusions, words written down which were not in the original passage. A better way to mark dictations, therefore, is to award students the marks for each section, one for each word in the correct version, and then subtract from that one mark for each mistake. Intrusions count as one mistake each. This could result in a minus score for some sections, in which case it is normal to award a zero.

There is a variation on the dictation test which has been used quite extensively in research on second language acquisition and in testing non-literate second language users. This is what is called a sentence elicitation task. It is basically the same as a dictation, except that students do not write down what they hear, but repeat it into a tape recorder during the pause after each section. This has shown itself to be a reasonably reliable method of testing general language ability (Gallimore & Tharp, 1981), but it suffers from the disadvantage of not being very suitable for the sort of large-scale testing which is necessary in Japan. It also seems to be unlikely to lead to a good washback effect on the classroom, as students will likely spend long periods of time just repeating sentences after a tape recording.<sup>5</sup>

Dictations are now firmly established in the second language teaching world as good tests of general second language ability, which are reasonably easy to make, not too difficult to mark and are generally

reliable and reasonably valid.<sup>6</sup> Dictations obviously require students to write, and listening scores may be influenced by second language writing ability in cases where students have widely different levels of written English. However, the writer has administered a number of English dictations to college students in Japan, and has found that most students have sufficient writing ability to handle dictations quite well. Dictations also seemed to cause far less student animosity than cloze tests. Furthermore, they seem to be considered quite acceptable academic tasks by the majority of teachers. However, if we want students to spend their classroom time listening to realistic communicative spoken language, then their washback effect will probably not be as positive as we would like. Students will likely practise listening to decontextualized written texts read out aloud, which is very different from listening to communicative spoken language. With this one very serious reservation, dictation does seem to be a possible candidate for inclusion in entrance examinations.

### *3.2 The Product Approach*

As we can now see, the main problem with testing the *process* of listening comprehension is that we don't really know enough about it. An obvious alternative approach is to test the *product* of listening. We can give students a passage to listen to and then see if they understood it. The drawback with this approach is that the product of the listening comprehension process is not easily available for inspection. It lies inside the student and we cannot record it, or take it home for grading. Assessing listening ability is an indirect process. We have to give students some task which we think is dependent on comprehending a piece of language, and then try to infer from that whether we think they understood the language or not. The task most commonly given to students is answering questions, but it could also be filling in a grid, or marking a place on a map or chart. Picture identification tasks are also sometimes used.

Naturally, it is important that completing the task is dependent on understanding the text. Care should be taken that task completion really does require comprehension of the text. It is surprising how often a good guess can be made at the right answer without even hearing the actual text (Preston, 1964; Connor & Read, 1978).

### *3.2.1 Intervening Variables*

There is, then, no such thing as a "pure" test of listening. We have to use indirect measures of listening, and when we choose a test format, it is important to realize that answering the questions will require other abilities apart from those we want to measure. This, of course, will confound the measurement of listening comprehension with other variables which we may not be interested in measuring at all. For instance, if the questions on an English second language listening test are presented to the student in written English, then ability to answer the questions will depend not only on second language listening ability, but also on second language reading ability.<sup>7</sup> A student with very slow reading may fail to answer many of the questions because he couldn't read fast enough, even though he may have understood the spoken text quite well.

One possible way around this problem is to give the question in oral form at the same time as the text is presented. The students listen to a passage, recorded on a tape, and then listen to the questions, which are also recorded on the tape.<sup>8</sup> This solves the problem of L2 reading ability contaminating listening tests, but introduces a different problem, namely that some students have better memories than others. Students with poor memories will obviously suffer. This might not be such a problem if the questions are rather short, and not such a strain on memory, but it could be a serious problem with multiple choice questions. Such tests also call for a high level of concentration from the test taker, as one short moment of inattention can result in the student not hearing the question and being unable to answer despite the fact that he understood the text quite well.

To some extent these problems can be avoided by asked open-ended questions, which would require less reading or less memorization. However, if the answers are required in the second language, then, of course, the listening comprehension test score will probably be contaminated by writing ability. One simple solution, at least in cases where all the students share the same first language, as in Japan, is to ask questions using the students' first language, and also allow them to answer in their first language. Although students obviously vary in their first language abilities, the difficulty of the Japanese used can be kept to such a basic level that it is unlikely to tax the ability of even the least proficient student. This avoids many of the problems mentioned above,

but moving backwards and forwards between two languages can cause problems, and some students find it irritating.

The choice of question form will depend largely on the purpose of the test, and the priorities of the test maker. Some teachers, for example, may be quite happy to have questions written in English, and thus have their listening tests contaminated by student reading abilities. After all, listening and reading are both important language skills, which are obviously closely related, and a test which measured both could well be a good test of general language comprehension. In the case of test instructions, though, the problem is rather more serious, as the penalty for misunderstanding one simple instruction can be the loss of many marks on the test. A penalty usually far out of proportion to the mistake made, which would suggest that, wherever possible, instructions should be in the students' first language.

The writer has used both multiple choice and open-ended questions, written in Japanese, on English listening tests and has found they work quite well if a little care is taken. On the open-ended questions, students were given the option of answering in either Japanese or English, which seemed the best solution, and results were generally satisfactory. Furthermore, using Japanese questions enables the test writer to ask questions which are more complex, and hence probe deeper into the students' understanding of the passage.

### *3.2.2 A Purpose for Listening*

In "real life," when we listen to anything, we virtually always listen with some purpose, even if the purpose is something as vague as general interest. If we want to replicate real-life listening, it is very important to give students a purpose for listening. This also relieves them from the necessity of trying to remember all the information in the passage, which is a very unnatural thing to do (and generally impossible), although it is quite common for students to try to do so in important testing situations. There are a number of ways the listening purpose can be set up, but perhaps the simplest is just to give students the questions before they hear the passage. A somewhat more interesting alternative is to give specific instructions about the listening situation and the purpose of their listening, a sort of listening role-play, and then allow students to select the relevant information for themselves.

### *3.2.3 Authenticity*

Another important consideration is the nature of the texts used. There is currently a very strong trend in language teaching to provide students with "authentic" language. While it is often difficult to be specific about exactly what we mean by authentic (Besse, 1981), the trend is an admirable one. Obviously, if we want our students to learn a language, we must give them realistic language to learn, and have them engage in realistic language activities with it. Of course, in one sense, the language learning classroom itself is not really an authentic situation, and the testing situation is even less an authentic communicative situation than the classroom. Whatever we try to get our students to do, however much we try to set up authentic situations, our students' prime concern will be to get high marks on the test. Nevertheless, given the tremendous washback effect of the test on the classroom, test makers owe it to teachers and students to use texts which are as authentic as possible, in order to encourage students to study authentic texts in the classroom.

### *3.2.4 Making Multiple Choice Tests*

In most standardized tests, such as the TOEFL or TOEIC, it is normal to use multiple choice questions to test comprehension. These are very convenient to score with modern high-speed marking machines, but making them, however, is not so easy. Not only do we have to write a sensible question and provide the correct answer, but we must think of three distractors, or alternative answers, which must look like attractive possibilities to those who did not understand the text, but which are clearly wrong to those who did. This is extremely difficult to do. Multiple choice items are very complex, and it is often difficult to tell how they are going to work by just looking at them. Therefore it is necessary to pre-test them on students similar to those who are to take the test, and then subject the results to statistical analysis in order to find out which items are good and which are not. Bad questions, or distractors which are not distracting any of the students, can then be rewritten; and it may often be necessary to reject or rewrite large numbers of the items. Using such a method, it is quite possible to produce tests which have very satisfactory statistical properties. However, pre-testing important entrance examination questions in

Japan is generally considered impossible, which suggests that test makers should avoid using multiple choice questions.

### *3.2.5 Open-ended Questions*

Of course it is possible to have comprehension questions without using a multiple choice format. One of the simplest ways is just to ask open-ended questions. Students then write the answer in their own words. The most obvious problem is that we then have to mark the answers, and that clearly involves deciding what is to be marked right and what wrong. Such judgments can be very difficult when the student has produced something which seems almost right, but which we feel is in some way inadequate. A related problem is how to deal with English answers that are grammatically incorrect. It seems reasonable to ignore small mistakes, but if the mistakes are such that they cause ambiguity, we have the problem of deciding what the student meant to say. It is necessary to get the criteria very clear, and ensure that all markers are applying the same standards. This problem can be avoided to some extent by using short-answer questions, in which students have to give very short answers of only one or two words. Such question types are ideal in the case of texts such as public announcements, or other information-style texts, where the main point of the message is clear and thus what answers are acceptable and what are not. Furthermore, the washback is likely to be very good, as such passages are a type of text we would like our students to study in the classroom. They are, after all, just the sort of thing they will have to understand if they go to an English-speaking country. Such short-answer questions are quite easy to make, students only spend a short time writing the answers, they are easy to mark, and answers can often be given in Japanese. In fact, they would seem to be ideal candidates for inclusion on college entrance examinations.

There are a lot of texts, though, which are not suitable for such short-answer questions. Sometimes teachers want to test more than just the ability to extract clearly stated information; they may also want to examine whether students have understood the main point of a passage, can produce a summary, or can make inferences about the events in the text. In such cases longer answers can produce far more information. However, the problem of deciding which answers are suitable and which are not becomes far more acute, as responses will vary on a

continuum from completely satisfactory to totally inadequate. The markers will often have to use a scale to rate the suitability of responses, and this is a far cry from simply marking items right or wrong.

Some sort of rating scale has to be developed, which need not be too complex, but which will certainly take some time and effort to produce. Exactly how this is made will depend on a number of factors, but basically it is necessary to decide what answers students are likely to produce, decide how these should be evaluated relative to each other, and decide how many marks to award for each. Having developed a scale to rate the answers, the job has only begun. It is then necessary to find out if the scale works on sample test items, after which it will probably be necessary to modify the scale a little. Next, raters must practice using the scale until they can all apply it consistently. This last step is very important. Students whose papers were marked by a generous rater have a far higher chance of passing the examination than those whose papers are marked by a stricter rater. Even the same rater tends to vary in the way he applies a rating scale over a period of time. The rater may get more severe, or less severe, or start paying more attention to one aspect of the scale than before. This variation in application of the rating scale is not only unfair, but can turn the examination into a worthless lottery.

### *3.2.6 Grids, Diagrams and Pictures*

Because of the problems associated with using questions in comprehension tests, some test makers have suggested trying to avoid these all together, by using diagrams or pictures (Heaton, 1975; Brown & Yule, 1983). Here again, though, there are a number of things to be considered and options to be weighed. First, while it is easy to use these to test simple or short samples of language, it is often not so easy to use them to test more complex language or extended discourse. Diagrams, and even more so, pictures, can take a considerable amount of preparation time, and often require drawing skills which many of us don't possess. Test makers who are good at drawing may enjoy making picture recognition tasks, in which students have to choose a picture to match a description they hear, but most of us are not very good at drawing and have to search around in old textbooks and such places to try to find pictures which we can use. Although picture recognition tasks may not look very serious or academic, they are used on the

TOEIC, which is a generally respected examination.

However, diagrams do offer possibilities of testing realistic language in a number of important situations. One obvious possibility is to give students a map and have them follow a conversation in which directions are given. Students follow the directions and then indicate on their answer sheet what the destination was. Another possibility is to give students an incomplete diagram, or a picture, and have them complete it by listening to a description of it. This sort of activity is very good for the language classroom, and can be enjoyable and extremely useful, but again, somehow, one wonders if they would be considered appropriate for entrance examinations. Perhaps a compromise could be found in filling in grids, which are something like a halfway point between short-answer questions and diagrams. Grids can be used for a host of activities, almost anything which involves timetables, or classifications (of objects or countries or blood types or whatever). They can be made such that almost no reading ability or memory is involved, and very little writing.

Almost all the activities mentioned in this section could lead to quite a good washback effect on the classroom, encouraging students to listen to railway announcements, weather forecasts and all sorts of other realistic and useful things. It would be very nice to see some of these tasks used in entrance examinations.

### 4. Conclusions

The first conclusion to note is that testing listening comprehension is far more complex an undertaking than it would seem to be at first sight. However, this applies not only to testing listening comprehension, but to testing any aspect of second language ability. Language use is a very complex phenomenon, about which far too little is known. The question test makers have to ask themselves is whether they are prepared to face up to the complexity of the issue or not. The writer believes very strongly that, as professionals, there is only one answer we can possibly give to such a question.

This article has discussed eleven test types. Table One lists the types, comments upon them, and estimates their likely washback effects.

In practical terms, the choice of test type will depend to some extent on the needs of the situation and the opinions of individual teachers. Of the various test types listed in Table One, cloze passages on summaries



of texts seem very suitable for entrance examinations and are obviously highly recommended, especially with narrative texts. Short-answer questions on realistic public announcements are another obvious choice and should be seriously considered. For expository tests, more complex open-ended questions can be used, or grid-filling activities where suitable. Grids or picture selection tasks are very suitable for descriptive texts. There is no reason why questions and answers should not be in Japanese. The writer feels every entrance examination could benefit from a question in which students hear directions on how to get to a destination and then find the destination on a map. Dictation is a possibility, or rather a temptation, because, although it is obviously a good testing technique, the likely washback effect argues against it.

The most important thing to bear in mind when making college entrance examinations in the Japanese situation is the washback effect on the classroom. If test makers continually ask themselves what sort of washback their test will lead to, there is every possibility of improving English language education in Japan.

## Testing Listening Comprehension

**Table One**  
**Summary of Listening Comprehension Tests**

Test	Comments	Washback
Noise tests	Utilizes little non-linguistic knowledge. Tape-recorded responses are impractical. Dictation is probably just as good, and more acceptable.	Very negative.
Ordinary listening cloze	Can often be done without listening. Students often just listen for individual words.	Probably not very positive.
Cloze on <i>summary</i> of passage	Students must understand whole passage. Can test inferencing and higher order skills. Summaries must be carefully produced, but summaries in Japanese are possible. Highly recommended.	Likely to be very positive indeed.
Dictation	Requires L2 writing ability. Well established testing technique. Easy to make and administer.	Rather negative.
Sentence elicitation	Same as dictation except that students record responses into a tape recorder. Rather impractical for large numbers. Good for non-literate students.	Rather negative.
Multiple choice comprehension questions	Can be machine-scored. Difficult to make, require pre-testing and item analysis.	Depends on questions made.
Short answer comprehension questions	Easy to make, and not too difficult to mark. Can be used with realistic communicative texts. Suitable for questions and answers in Japanese. Highly recommended.	Very positive if used on realistic texts.
Open-ended longer questions	Easy to make, but usually require a rating scale to mark them. This takes time and coordination between raters. Can be used to test understanding in depth.	Quite positive if used on realistic texts.
Picture recognition tasks	Requires a lot of time or drawing skill to make. Take up a lot of space on test paper. Don't look very academic.	Good if natural English and realistic tasks are used.
Diagram completion tasks	Easier to make than pictures. Can be used to test natural language use. Following street directions is a very good communicative activity easily incorporated in any test. Don't look very academic.	Good if natural English and realistic tasks are used.
Grid completion	Easy to make. Can be used to test natural language. Endless possibilities, especially with timetables and classifications.	Good if natural English and realistic tasks are used.

### Notes

1. It is a basic truism of educational measurement that if a test is too difficult for the students on which it is used, it will actually give very little information regarding their ability.
2. The writer does not feel a need to distinguish between the terms acquisition and learning. They will be used interchangeably in this paper.
3. I would like to take this opportunity to offer thanks to the teachers at Osaka Meijo Women's College who kindly cooperated in this study.
4. Cronbach Alpha makes the assumption that the items in the test are independent of each other. If they are not, the reliability estimate appears higher than it should be. It is possible that the items in this test, as with all cloze tests, don't meet this requirement in all respects and thus the reliability could be a little lower.
5. This might be acceptable to those who feel that the Audio Lingual Method is a good way to learn languages, but of course, such a view is now very rare among applied linguists.
6. It should be noted that there is no such thing as a testing method which automatically produces reliable and valid tests. Unfortunately, it is necessary to validate each new test, and even each new use of an old test.
7. A look at TOEFL practice listening tests will show just how much reading is actually involved in some listening tests.
8. Both the TOEIC listening section, and the JACET-COLTD Listening Comprehension Test have questions recorded on the tape.

### References

- Alderson, J. C., & Urquhart, A. H. (eds.). (1984). *Reading in a foreign language*. London: Longman.
- Besse, H. (1981). The pedagogic authenticity of a text. In *ELT Documents special: The teaching of listening comprehension*. London: The British Council.
- Brown, G., & Yule, G. (1983). *Teaching the spoken language*. Cambridge: Cambridge University Press.
- Bruce D. J. (1985). The effects of listeners' anticipations on the intelligibility of heard speech. *Language and Speech* 1, 79-97.
- Buck, G. (in press). *Written pronunciation tests: Do they work?* *ELT Journal*.
- Carrell, P. L. (1987). Content and formal schemata. *TESOL Quarterly* 21, 461-481.
- Cohen, A. (1977). Redundancy as a tool in measuring listening comprehension. In R. Dirven (ed.), *Horverstandnis im Fremdsprachenunterricht. Listening comprehension in foreign language teaching*. Kronberg/Ts.: Scriptor.
- tests. *Language Learning* 28 (1), 149-157.
- Devine, J., Carrell, P. L., & Eskey, D. E. (eds.). (1987). *Research in reading in a foreign language*. Washington, D.C.: TESOL.
- Dirven, R., & Oakeshott-Taylor, J. (1984). Listening comprehension (Part I). *Language Teaching* 17, 326-342.
- Dirven, R., & Oakeshott-Taylor, J. (1985). Listening comprehension (Part II), *Language Teaching* 18, 2-20.

## Testing Listening Comprehension

- Faerch, C., & Kasper, G. (1986). The role of comprehension in second language learning. *Applied Linguistics* 7, 257-274.
- Gallimore, R., & Tharp, R. G. (1981). The interpretation of elicited sentence imitation in a standardized context. *Language Learning* 31 (2), 370-392.
- Goodman, K. (1970). Reading: A psycholinguistic guessing game. In H. Singer & R. B. Ruddell (eds.), *Theoretical models and processes of reading*. Newark, Del.: International Reading Association.
- Gradman, H. L., & Spolsky, B. (1975). Reduced redundancy testing: A progress report. In R. L. Jones & B. Spolsky (eds.), *Testing language proficiency*. Arlington, Va.: Center for Applied Linguistics.
- Grosjean, F. (1980). Spoken word recognition and the gating paradigm. *Perception and Psychophysics* 28, 267-283.
- Heaton, J. B. (1975). *Writing English language tests*. London: Longman.
- Henning, G., Gary, N., & Gary, J. (1983). Listening recall: A listening comprehension test for low proficiency learners. *System* 11, 287-293.
- Johansson, S. (1973). An evaluation of the noise test: A method for testing overall second language proficiency by perception under masking noise. *IRAL* 11, 107-131.
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon Press.
- Lado, R. (1961). *Language testing*. London: Longman.
- Leech, G. N. (1983). *Principles of pragmatics*. London: Longman.
- Marslen-Wilson, W. D. (1980). Speech understanding as a psychological process. In J. C. Simon (ed.), *Spoken language generation and understanding*. Dordrecht: Reibel.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review* 63 (2), 81-97.
- Oakeshott-Taylor, A. (1977). Dictation as a test of listening comprehension. In R. Dirven (ed.), *Horverstandnis im Fremdsprachenunterricht. Listening comprehension in foreign language teaching*. Kronberg/Ts.: Scriptor.
- Oller, W. J. (1979). *Language tests at school*. London: Longman.
- Preston, R. C. (1964). Ability of students to identify correct responses before reading. *The Journal of Educational Research* 58 (4), 181-183.
- Richards, J. C., & Rodgers, T. S. (1986). *Approaches and methods in language teaching*. Cambridge: Cambridge University Press.
- Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (eds.), *Theoretical issues in reading comprehension*. Hillsdale, N.J.: Lawrence Erlbaum.
- Rumelhart, D. E., & Ortony, A. (1977). The representation of knowledge in memory. In R. C. Anderson, R. J. Spiro, & W. E. Montague (eds.), *Schooling and the acquisition of knowledge*. Hillsdale, N.J.: Lawrence Erlbaum.
- Savignon, S. J. (1983). *Communicative competence: Theory and practice*. Reading, Mass.: Addison-Wesley.
- Schank, R. C., & Abelson, R. P. (1977). Scripts, plans, and knowledge. In P. N. Johnson-Laird & P. C. Wason (eds.), *Thinking: Readings in cognitive science*. Cambridge: Cambridge University Press.
- Smith, F. (1985). *Reading*. Cambridge: Cambridge University Press.
- Spolsky, B. (1971). Reduced redundancy as a language testing tool. In G. E. Perren & J. L. M. Trim (eds.), *Applications of linguistics*. Cambridge: Cambridge University Press.
- Spolsky, B. (1975). Language testing: Art or science? Paper presented Aug. 27, 1975 at the Fourth AILA World Congress, Stuttgart, Germany.



# The Modern Language Journal

Founded: 1916

Editor: David P. Benseler  
Dept. of German  
Ohio State University  
Columbus, OH 43210

... is recognized throughout the world as the outstanding foreign language/pedagogical research journal in the United States. This popular journal was founded in 1916 by the National Federation of Modern Language Teachers Associations. Topics covered in MLJ include teaching strategies, bilingualism, applied linguistics, teaching of literature, study abroad, innovative foreign language programs, and developments in curriculum, teaching materials, and testing and evaluation.

ISSN 0026-7902

**Published quarterly at The University of Wisconsin Press**

Subscribe now, or recommend a subscription to your library.  
A detailed brochure will be sent upon request.

#### RATES

Individuals: \$13/yr.  
(must prepay)  
Institutions: \$30/yr.  
Foreign subscribers add \$3/yr.  
for regular postage, or \$12.50/yr.  
for Airmail delivery.

#### REPLY TO

Journal Division  
The University of Wisconsin  
Press  
114 North Murray Street  
Madison, Wisconsin 53715  
USA  
(608) 262-4952

---



# New books for teachers

## Understanding Research in Second Language Learning

A teacher's guide to statistics and research design

**James Dean Brown**

Designed specifically for language teachers with no previous background in statistics, this book focuses on the skills and processes necessary for understanding and critically reading statistical research in language learning. Brown explains the basic terms of statistics; the structure and organisation of statistical research reports; the system of statistical logic; and how to decipher tables, charts and graphs.

*'This book does an excellent job of filling a major gap in the publications available in our field.'*

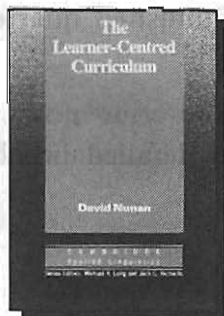
Lyle F. Bachman, University of Illinois at Urbana-Champaign

## *New Directions in Language Teaching* **The Learner-Centred Curriculum**

**David Nunan**

Traditionally, the curriculum has been regarded as the statement of what *should* be done in a course of study. This book takes as its starting point what *is* done by language teachers in their classes. The concept of a negotiated model is developed, in which the curriculum is arrived at collaboratively between teachers and learners.

The author looks at the curriculum from the teacher's perspective, and reports on what teachers focus on in the planning, implementation and evaluation of language courses.



## *Cambridge Applied Linguistics*

For further information on all Cambridge ELT publications, please contact  
Steven Maginn, Cambridge ELT Office,  
c/o U.P.S. Ltd., Kenkyu-sha Bldg.,  
9 Kanda Surugadai 2-chome, Chiyoda-ku,  
Tokyo 101. Tel: (03) 295-5875.

**CAMBRIDGE UNIVERSITY PRESS**