CHAPTER ONE

....

Why test language for specific purposes?

Introduction

Testing language for specific purposes (LSP) refers to that branch of language testing in which the test content and test methods are derived from an analysis of a specific language use situation, such as Spanish for Business, Japanese for Tour Guides, Italian for Language Teachers, or English for Air Traffic Control. LSP tests are usually contrasted with general purpose language tests, in which purpose is more broadly defined, as in the *Test of English as a Foreign Language* (*TOEFL*) (Educational Testing Service 1965). As you will see, it is important to note that tests are not either general purpose or specific purpose – all tests are developed for some purpose – but that there is a continuum of specificity from very general to very specific, and a given test may fall at any point on the continuum. I will argue later in this chapter that LSP testing is a special case of communicative language testing, since both are based on a theoretical construct of contextualized communicative language ability, and that LSP tests are no different in terms of the qualities of good testing practice from other types of language tests.

I should note that, over the years since its beginnings, specific purpose language testing has been criticized on a number of grounds: specific purpose language proficiency is really just general purpose language proficiency with technical vocabulary thrown in; we don't need specific purpose tests since, if we test general language knowledge, specific uses will take care of themselves; specific purpose

1

language tests are unreliable and invalid since subject knowledge interferes with the measurement of language knowledge; there is no theoretical justification for specific purpose language testing; and specific purpose language testing is impossible anyway, since the logical end of specificity is a test for one person at one point in time. In this book, I intend to refute these and other arguments in favor of the view that specific purpose language tests are indeed necessary, reliable, valid, and theoretically well-motivated.

Typically, LSP tests have been construed as those involving language for academic purposes and for occupational or professional purposes. Readers may wish to have a look at the following publications for further information on the field of language for specific purposes, of which LSP testing is certainly a part: Swales (1985) for a discussion of the development of the field, and Dudley-Evans and St John (1998) for a discussion of current developments. I will focus on two aspects of LSP testing that may be said to distinguish it from more general purpose language testing: **authenticity of task** and the **interaction between language knowledge and specific purpose content knowledge**. Authenticity of task means that the LSP test tasks should share critical features of tasks in the target language use situation of interest to the test takers. The intent of linking the test tasks to non-test tasks in this way is to increase the likelihood that the test taker will carry out the test task in the same way as the task would be carried out in the actual target situation. The interaction between language knowledge and content, or background, knowledge is perhaps the clearest defining feature of LSP testing, for in more general purpose language testing, the factor of background knowledge is usually seen as a confounding variable, contributing to measurement error and to be minimized as much as possible. In LSP testing, on the other hand, as you will see in Chapter 2, background knowledge is a necessary, integral part of the concept of specific purpose language ability.

LSP testing, like LSP teaching, has a relatively short history. A case could be made for the beginning of LSP testing as early as 1913, with the establishment of the University of Cambridge Local Examinations Syndicate's (UCLES) *Certificate of Proficiency in English*, a test designed for prospective English teachers to demonstrate their proficiency in the language (University of Cambridge Local Examinations Syndicate 1995). Another candidate for the title of first LSP test might be the College Entrance Examination Board's *English Competence* examina-

tion in the US, a test for international applicants to US colleges and universities introduced in 1930 (Spolsky 1995). Both of these tests have clearly defined purposes related to vocational and academic English, respectively, and thus in a sense qualify as examples of LSP tests. However, as I mentioned above, LSP testing requires first, an analysis of a target language use situation, from which characteristics of test content and method are derived, as well as an interaction between language knowledge and specific purpose content knowledge. Clearly, not all examples of what we call LSP tests manage to meet these criteria completely, but I will argue in this book that a theory of LSP testing establishes these two characteristics as fundamental goals. The UCLES and the College Board tests were not developed on the basis of analyses of language teaching or academic situations, nor did the tasks on the tests bear much relationship to the kinds of tasks required of either teachers or students (except when taking language tests!).

So, when might we say that true LSP testing began? A strong candidate is the *Temporary Registration Assessment Board* (*TRAB*) examination, a test introduced in 1975 by the British General Medical Council for the purpose of evaluating the professional and language abilities of physicians trained outside the UK applying for temporary registration to practice medicine in Britain (Rea-Dickens 1987). The examination consisted of an assessment of both professional competence and ability to communicate in English. The language component comprised a taped listening test, a written essay, and an oral interview in which both professional knowledge and language ability were assessed. The *TRAB* language component was based on an analysis of the language, both spoken and written, actually used by physicians, nurses, and patients in British hospitals. As I have discussed, this analytical approach is a critical feature of LSP test development. In addition, the language testing specialists who developed the language component of the *TRAB* test were not solely responsible for its development, but worked together with medical experts in constructing the tests. This is an important aspect of specific purpose test development. As Rea-Dickins (1987) put it in discussing the *TRAB* development process, collaboration with practitioners in the specialist area 'would seem to be a pre-requisite for the design of a ''special purposes'' test as the domains incorporated within the specialist area go beyond those in which the linguist – independently – is competent to make judgements' (p. 196). Thirdly, the *TRAB* developers attempted

to promote the engagement of the test takers' language ability and background knowledge in the test tasks by providing appropriate and rich contextual features in the test material. For example, in the writing tasks, the test takers were presented with authentic information about a patient's case history, and the tasks were linked systematically to the problems presented. Typical writing tasks included the following:

> Write a letter to Dr Jones summarising the case and giving your recommendations for Mr Brown's after-care.
>
> Complete the x-ray request card for this examination.
>
> When the patient is admitted to hospital, what written instructions would you leave the night nurse in charge of the ward regarding management?
>
> <div align="right">Rea-Dickins (1987: 195)</div>

We can see in this early example of an LSP test the embodiment of the critical features of LSP test development: analysis of the target language use situation, authenticity of task, and interaction between language and content knowledge. The *TRAB* was later revised (its name changed to *PLAB – Professional and Linguistic Assessment Board*), and is at present no longer in use, but it stands as a worthy prototype of the art of LSP test development. (Readers might also want to note another early LSP test, the *English Language Teaching Development Unit* [*ELTDU*] test, introduced in 1976 as an assessment of vocational English. See North 1994 for information.)

You might reasonably ask the question, however, as to why LSP testing is necessary, or even desirable. To consider this issue, let us imagine a typical language testing situation. As in all good language testing projects, LSP test development begins with a problem to be solved.

A problem

Suppose we want to determine whether people involved in international trade know English well enough to conduct their business. In such a situation, we might reasonably decide to devise a test of English for international business purposes. We would begin our task as test developers by interviewing experienced business people, as well as company supervisors, heads of international divisions, and an

assortment of middle level managers who typically deal with interna-
tional colleagues. We might observe actual negotiating sessions and
business meetings, and tape record participants' use of English in the
various situations they find themselves in: large meetings, one-on-
one discussions in offices, individual and conference telephone calls,
the ubiquitous business lunch and other business-related social occa-
sions, and so on. Our goal would be to describe the situations in
which international business people conduct their work, and the
characteristics of the language they use and of the tasks they must
perform in English.

We would need to make some decisions about the scope and
content of our test. For example, how important is it to test ability to
communicate about food or travel? Should we require the test candi-
dates to demonstrate knowledge of their field of business as well as
their abilities in English? Such decisions would have to be made in
consultation with the sponsors of the test, for their purposes in
wishing to give the test – and their willingness to pay for a longer and
more varied test! – will help determine what aspects of the milieu of
international business we will include in our test. Eventually,
however, we would be in a position to produce **test specifications**, a
blueprint of the test we intend to develop, including a statement of
the purpose of the test, a description of what it is we intend to
measure, a description of the contexts and tasks we intend to include
in the test (based on our analysis of the features of the international
business domain), details of how the test will be scored, and an
indication of how scores on the test should be interpreted.

On the basis of these specifications, we would then actually
produce test tasks and assemble a specific purpose test of English for
international business. After trying the new test out, perhaps by
giving it to a group of business people, and revising it, we would offer
it to our target group of prospective international traders. We would
interpret their performance on our test as evidence that they could, or
could not, use English well enough to succeed in the tasks required of
them in the marketplace.

Why bother?

But why go to all the trouble of devising a new test? Why spend the
time, effort, and money to interview people, describe the language

tasks of international business, devise the test, and pilot and revise it? Why not just turn to an existing test of English language ability, one such as the Educational Testing Service's *Test of English as a Foreign Language* (*TOEFL*), or the Cambridge University Local Examinations Syndicate's *Certificate of Proficiency in English* (*CPE*)? These, after all, are well-known international tests, with well-known measurement properties. The *TOEFL* is a multiple-choice test of listening, structure, reading comprehension, and writing, and is often taken by people who wish to demonstrate English language ability for international communication. *TOEFL* candidates can opt to take a speaking test as well, to further demonstrate their ability to use English. The *CPE* is a general test of English reading, writing, structure, listening, and speaking, and is used by many businesses to certify the English language skills of their employees, in addition to its main purpose for university admissions. So, why not use an existing, general purpose language test for our international business candidates?

Reason 1: language performances vary with context

One reason is that researchers are pretty much in agreement that language performances vary with both context and test task, and therefore our interpretations of a test taker's language ability must vary from performance to performance. For example, if we give test takers a reading test based on a passage about square-rigged sailing ships, followed by one based on a passage about micro-chips in computers, they will probably perform somewhat differently on the two tests, particularly if they are studying computer engineering! However, as you will see, it is not enough merely to give test takers topics relevant to the field they are studying or working in: the material the test is based on must engage test takers in a task in which both language ability and knowledge of the field interact with the test content in a way which is similar to the target language use situation. The test task, in other words, must be **authentic** for it to represent a specific purpose field in any measurable way. I will discuss the nature of authenticity in more detail below, but for now let us agree that LSP testing requires the use of field specific content in tasks which might plausibly be carried out in those fields. Returning to our business English example, it would not be enough, in this view, to provide test takers with listening texts about the work of international commerce,

but rather it would be necessary to provide test tasks that share similar characteristics with the tasks that international traders actually perform in their work, both in the processing of information and in responding to it. Thus we must keep in mind that an important reason for using specific purpose measures is that if we wish to interpret a person's test performance as evidence of language ability in a specific language use situation, we must engage the test taker in tasks which are authentically representative of that situation.

There is quite a bit of research which suggests that this interaction between the test taker's language ability and specific purpose content knowledge and the test task is a necessary condition in LSP tests. It has been found, for example, that when test takers have some prior knowledge of the topic of a reading passage, they have an advantage in responding to comprehension questions based on that passage. This suggests that there may be no such animal as a pure language test. Measures of language ability are always colored by such factors as background knowledge and test method. It has also been found, however, that the advantage due to specific purpose content knowledge may be quite negligible unless the passage and tasks are sufficiently specific to engage the test takers in authentic language use. I will discuss evidence for this claim in some detail in Chapter 2.

Reason 2: specific purpose language is precise

A second reason for preferring LSP tests over more general ones is that technical language – that used in any academic, professional or vocational field, including cooking, law, physics, chemistry, air traffic control, scuba diving, religion, stamp collecting, or language teaching – has specific characteristics that people who work in the field must control. What we often refer to as jargon or even gobbledygook has a specific communicative function within that field, namely **precision**. There are lexical, semantic, syntactic, and even phonological characteristics of language peculiar to any field, and these characteristics allow for people in that field to speak and write more precisely about aspects of the field that outsiders sometimes find impenetrable. It is this precision that is a major focus of specific purpose language use and is a major factor arguing in favor of specific purpose language tests. A classic example of the need for precise, specific purpose language comes from the field of law. We frequently deplore what we

call legalese, the arcane lexis, the convoluted syntax, the use of Latin terminology, and the interminable cross-references to previous laws and cases in legal texts. Yet, legal language was purposefully developed and is used dynamically by members of the legal profession to communicate among themselves the precise meaning of the law. A good example can be found on the back of any airline ticket:

Conditions of Contract

1   As used in this contract, 'ticket' means this passenger ticket and baggage check, of which these conditions and the notices form part, 'carriage' is equivalent to 'transportation,' 'carrier' means all air carriers that carry or undertake to carry the passenger or his baggage hereunder or perform any other service incidental to such air carriage, 'WARSAW CONVENTION' means the Convention for the Unification of Certain Rules Relating to International Carriage by Air signed at Warsaw, 12th October 1929, or that Convention as amended at The Hague, 28th September 1955, whichever may be applicable.   Ref. 1293 ATB (REV. 6–89)

This statement, not among the most opaque of legalese, but nevertheless quite recognizable as such, was clearly written not by the airline's public relations officer for the traveler who bought the ticket, but rather by lawyers for other lawyers, and is a good example of the legal profession's demand for precision in language. If, for whatever reason, we wanted to measure a lawyer's control of English to conduct the business of law, it would not seem to be sufficient to use texts and tasks which were not specific to the legal profession. There may be perfectly good reasons to include language and tasks not so strictly related to the legal register in the test, but certainly if our goal is to measure a test taker's ability to use language within a specific vocation, profession, or academic field, and that is the focus of this book, then specific purpose texts and tasks will be needed.

How are specific purpose language tests related to other types of language tests?

Speaking of precision, it is, of course, necessary to be more precise about the nature of specific purpose language tests than I have been so far. For the moment, let us agree to define our object of interest as tests which attempt to measure language ability for specific

vocational, professional, and academic purposes. I will suggest a more precise definition later in this chapter, but before we can arrive at a useful definition of specific purpose language testing, we need to discuss a number of related concepts in language testing that form the background to LSP testing. These include **communicative testing**, **general proficiency testing**, **criterion-referenced testing**, and the notion of **authenticity**.

Communicative tests

Particularly since the publication in 1978 of Widdowson's book, *Teaching language as communication*, and in 1980 of Canale and Swain's paper, 'Theoretical bases of communicative approaches to second language teaching and testing,' the related fields of language pedagogy and language assessment have been characterized by the communicative paradigm, the communicative approach, and communicative language teaching. But even before the publication of Hymes's (1972) classic paper, 'On communicative competence,' which provided much of the impetus for the communicative approach, language testers were discussing 'productive communication testing' (Upshur 1971), and teachers and testers have been fascinated with the notion for over a quarter of a century now. As you will see below, specific purpose language tests are by definition communicative. Indeed, Sajavaara (1992), in a discussion of LSP test design, assumes from the outset that 'It is impossible to distinguish LSP testing theoretically from communicative language testing' (p. 123).

In his book *Communicative language testing*, Weir defines his topic as follows:

> In testing communicative language ability we are evaluating samples of performance, in certain specific contexts of use, created under particular test constraints, for what they can tell us about a candidate's communicative capacity or language ability.
>
> Weir (1990: 7)

In his definition, Weir employs a number of key terms: **communicative language ability**, **specific contexts of use**, **test constraints**, and **capacity**. Since specific purpose language testing involves all these concepts, we will conceive of it as a special case of communicative language testing. The first of Weir's terms, communicative language

ability (CLA), was introduced by Bachman as a framework for describing language knowledge and the capacity for implementing it 'in appropriate, contextualized communicative language use' (Bachman 1990: 84). This leads us to Weir's second key term, specific contexts of use, requiring us to take account of the many features of context that influence communication, features such as the physical and temporal setting, the role(s) of the test taker and the interlocutor(s)/audience, the purposes of the communication, the topic and content of the message, its tone and manner, and the channels, codes, and genres being employed (cf. Hymes 1974). The third key term in Weir's definition, test constraints, reminds us that the methods we employ in eliciting a language performance will influence the nature of the performance and thus the interpretations we might make on the basis of it. Tests are, after all, contrived language use events, and even the most cleverly contrived test tasks limit to some degree the generalizability of our interpretations concerning the test takers' specific purpose language abilities.

Finally, Weir refers to capacity, a term employed by Widdowson (1983), as 'the ability to use knowledge of language as a resource for the creation of meaning' (p. 25), and is intended to be understood from the perspective of the language user rather than that of the language analyst (or, indeed, the language tester). In this book, I will use the term **communicative language ability** (and later, **specific purpose language ability**) to capture the notion of capacity as Weir and Widdowson use the term. The point that is crucial in the testing of language ability in specific purpose contexts is understanding that ability from the perspective of the language user. That is, not only are we interested in measuring communicative language ability rather than language performances *per se*, but we are called, in LSP testing, to interpret test performance from the point of view of language users in the specific purpose situation of interest. Thus, specific purpose language testing, as Widdowson points out with regard to specific purpose language teaching, is essentially an exercise rooted in an understanding of human activity from the point of view of the participants in the activity. In this regard, in Chapter 2, I will explore the concepts of grounded ethnography and indigenous assessment, as useful approaches for understanding the nature of LSP test performance from the point of view of the language users.