

Unit A2

Measuring bilingualism

What does it mean to know a language?

What does language knowledge consist of, and how do we measure it?

Should a bilingual's two languages be measured against the single language of a monolingual?

How do we measure the language knowledge of native speakers?

As you can probably infer from the discussion of the complex nature of defining bilingualism in Unit A1, the measurement of bilingual language skill is far from simple. We have seen that the range of what can constitute a definition of bilingualism falls anywhere along a continuum from full competence in one language plus limited knowledge of a second, to apparent full competence in all skills in two or more languages. In addition to this, as we saw, we need also to be aware of the importance of identifying the range of variables which impact on bilingual language use – such factors as the age of acquisition, the various domains of language use (which may be different for both languages), and the socio-cultural contexts in which the languages are used.

In this unit, we will begin with a discussion of whose bilingual skills we are measuring. Not everyone comes to bilingualism by the same route, and, as we will see, this has implications for both the kinds of measures we use and how we think about language knowledge. The first group we will consider are those people who for some reason choose to learn another language and thus move towards bilingualism. We will examine the kinds of testing instruments that are used for this group and this will introduce a number of concepts which are central to the understanding of the measurement of language.

Next, we will look at issues in testing linguistic minorities living in situations where their home language is different from that of the wider community. In particular, we will discuss the importance of using appropriate testing instruments both for educational assessments and for the identification of language difficulties (or specific language impairment) in children, and introduce the ways in which bilingualism is measured by researchers. As we will see in more detail in Unit A4, researchers need to assess the bilingualism of individuals in order to further our understanding of the range of effects of bilingualism on individuals. In the next

few sections, we will focus in particular on the assessment of children's language since this has been a major area of study.

WHOM ARE WE TESTING?

Tests of all kinds are generally developed for specific populations. Because of the very varied nature of bilingualism and the myriad different ways that people acquire some degree of bilingualism, it is helpful to delineate the circumstances under which we assess the language of people who are bilingual. Very roughly, we can consider classifying bilingual speakers into three different groups:

- developing bilinguals
- stable bilinguals
- attriting bilinguals.

Stable bilinguals are generally people for whom bilingualism is a way of life, and for whom the daily use of two languages is the norm rather than the exception. They may be either *elective* bilinguals or *circumstantial* bilinguals, although they are probably more likely to come from the latter group. They may use both languages in different contexts – for example, they may speak Language A at home and Language B at work, or they may use both languages in a variety of contexts. Alternatively, they may live in a multilingual community where two or more languages are used routinely.

Attriting bilinguals consist of those individuals who are, for some reason, suffering from some aspect of language attrition. This may result from lack of contact and lack of use of one of their languages (and this is often the case with elective bilinguals), or it may be the result of pathological factors either associated with age or as a consequence of some kind of accident. This general area of language attrition will be dealt with in more depth in Unit A5.

For the purposes of this unit, we will concentrate on the issues that surround the measurement of the language of what we term *developing* bilinguals.



Task A2.1

- Drawing on the people you know, try to think of some examples of people who are:
 - developing bilinguals
 - stable bilinguals
 - attriting bilinguals.
- Is it possible to be a member of more than one of these groups?

Developing bilinguals

We focus on this group because it is the largest and the most widely assessed group. As we discussed briefly in Unit A1, Valdes and Figueroa (1994) divide this group into two: *elective bilinguals* and *circumstantial bilinguals*. This is a useful division, and one which we will use to structure this unit. Each group has quite different characteristics, as shown in Table A2.1 (adapted from Valdes and Figueroa (1994: 13–14).

Table A2.1 Characteristics and examples of elective and circumstantial bilinguals

<i>Elective bilinguals</i>	<i>Circumstantial bilinguals</i>
Characteristic of individuals	Characteristic of groups
Choose to learn another language	Second language required to meet needs of new circumstances
Communicative opportunities usually sought artificially (e.g. in classroom)	Communicative needs may relate to survival, or success; communicative needs will vary across individuals
First language will usually remain the dominant language	Two languages will play a complementary role and the stronger language may vary depending on the domain
<i>Examples of elective bilinguals</i>	<i>Examples of circumstantial bilinguals</i>
A child raised with a French-speaking mother and Italian-speaking father in an English-speaking environment	Children raised in families where two languages are spoken both inside and outside the home
A Japanese student who has learned English in order to study for a Master of Arts degree in Australia	Immigrant groups who have moved to a country where another language is spoken
An American man who learns Russian because he has married a Russian woman and moved to Russia	Indigenous groups living in countries which have been colonized
A diplomat who learns Mandarin Chinese for her job	Groups whose first language is different from the prestige language of the surrounding community. (A prestige language is one which has higher status in the community than other languages spoken, usually because it is the language of education and government.)

As we saw in Unit A1, the major difference between elective bilinguals and circumstantial bilinguals is that, for the former, there is generally some element of choice about learning a second language, while the latter are obliged to learn another

language as a result of the circumstances in which they find themselves. Although these circumstances may be chosen, as in the case of choosing to emigrate to another country, they are frequently not a result of choice – for example, people forced to flee their country of origin as a result of war or famine, or indigenous people in countries whose language is not used as a result of colonization.

The implications for the assessment of bilingualism turn out, in fact, to be quite different for these two different groups. While these two groups will be used to structure our discussion of bilingual assessment, it is important to point out that although this distinction is a useful one as a framework for the discussion of bilingual assessment, it is not necessarily as clear-cut as Table 2.1 above implies.



Task A2.2

- Try to think of some examples of people who are 'elective' bilinguals and people who are 'circumstantial' bilinguals.
- Can an individual move from one group to the other?
- Can you be a member of both groups at the same time? How? Try to give an example of a person who would be considered to be both an elective bilingual and a circumstantial bilingual (assuming that this person is at least trilingual!).
- Would you yourself qualify for membership of one of the groups? Which one?

Assessing bilingual proficiency

The assessment of bilingual language proficiency is difficult in part because we are immediately confronted with the question of 'what is a bilingual', or, as Bialystok (2001a: 10) puts it, '*When is enough enough?*':

The problem of knowing who is bilingual conceals a more basic question: how much is enough? Who among us does not know pieces of some other language – words or phrases, perhaps a rule or two, and some social routines for greeting, toasting drinks, or asking directions? These fragments hardly count as competence in the language, but how much more is required before some implicit threshold is reached? Accepting the standard assumption that no bilingual is ever equally competent in both languages, how much language is needed before we agree that a person is bilingual?

The answer depends on how we define language proficiency. We talk about language as though it had concrete existence and could be measured by scientific instruments. We describe the acquisition of language as though we move irrevocably from a state of innocence to one of mastery along a predictable path. We identify language impairment, language delay, and language precocity without ever specifying the standard against which these

cases are to be judged. We use 'language' in research designs as both a dependent and an independent variable, choosing fragments to serve as stimuli but concluding truths that define the domain. But what is the norm for language competence? What do we mean by language proficiency? What are its components and what is the range of acceptable variation? Although these questions may seem to be prior to any use of language as a research instrument or conclusion about language ability in individuals, they rarely if ever are explicitly addressed.

What constitutes language proficiency is a difficult question. When we want to measure the language of bilinguals, we are often (though not always) concerned not only with measuring one language but with measuring both languages, and with measuring the interrelationships between these two languages. These issues will be discussed in more detail below.

We have used the distinction between elective and circumstantial bilinguals to help us respond to Bialystok's question of when 'enough is enough'. By drawing this distinction between the two groups, we will argue that for elective bilinguals we almost never assess both languages. Rather, we assume the speaker's proficiency in their first language, and assess only their proficiency in their second. For circumstantial bilinguals, however, the situation is rather different. This is because often we *should* be concerned with assessing both languages because circumstantial bilinguals use language in different contexts and to assess only one may not give us a full picture of the individual's overall language ability. Mills (1995: 144–145) provides a good example illustrating that children will not necessarily know the word for a concept in both their languages:

I encountered Kuldip (age six) who was a Panjabi-speaker and had been in the UK for six months. It became evident Kuldip did not know the names of the shapes 'square' and 'rectangle' in Panjabi, but *did* know them in English.

As we will see, there are often factors which make it either difficult or impossible to appropriately assess a child in both languages. Equally, while we can recognize that we *should* be assessing the interrelationship between the two languages, this is not always an easy prospect.

Task A2.3



- A child who goes to school at five has considerable linguistic knowledge, and, if the child is bilingual, the child has considerable linguistic knowledge in both languages. Think about the language knowledge children develop during their primary or elementary school years. Can you identify some areas in which the child might develop linguistic knowledge in the school language, but not the home language?

ASSESSING THE BILINGUAL PROFICIENCY OF ELECTIVE BILINGUALS

In the examples of elective bilinguals given in Table A2.1 (with the possible exception of the first situation of the child growing up in a simultaneously bilingual environment), it is clear that each of the individuals mentioned is a native speaker of one language, and is learning another language as a second language. In each case, the individual has elected to learn the other language and is motivated to learn it. For this group of individuals (i.e. elective bilinguals) we can almost always identify quite easily the language of which the individual is a native speaker. However, we generally do not seek to assess the proficiency of their native language. As we will see below, this marks a substantial difference in the way in which we think about assessing elective bilinguals compared with circumstantial bilinguals. Thus, when we are assessing the language of an elective bilingual, we are almost always concerned with assessing their second language: the role of their first language (the one of which they are a native speaker) does not generally come into consideration except inasmuch as it might impact on their acquisition of the second language. This is because we tend to accept native speakers as proficient speakers of the language even though, as native speakers, they may have greater or lesser vocabularies, and may each have slightly different grammars (Davies 1991b; see also Unit A1). Certainly not all native speakers speak identically, but when we talk about native speakers we tend to think about an idealized native speaker who demonstrates full competence in the language. In the discussion below, then, we focus on what we need to measure when we think about measuring the second-language proficiency of those people who, largely through choice, elect to learn another language.

Measuring language

The measurement of any phenomenon requires that initially we have a clear definition of the bounds of the phenomenon that we want to measure. With physical objects in the world this is considerably less problematic than with psychological constructs because physical entities generally (though certainly not always) are relatively clearly defined – we know where they begin and where they end. In order to be able to measure the object, in addition to knowing where it begins and ends, we need to have some kind of gauge that is meaningful, and with which we can expect to be able to make reliable evaluations. If we decide to measure the length of a table, we will probably use a tape measure of some kind. The scale of the tape measure may be either in centimetres or in inches. When we look at the table we can see its length, and we know that we need to measure from one end to the other, and note the length. Obviously, the measurement of much larger physical objects, such as the height of buildings, or much smaller physical objects, such as atoms, poses more complex problems. However, they are more complex problems because of their size, either their largeness or their smallness, which may mean that instruments need to be developed specifically, or that mathematical procedures, such as triangulation, will be required to complete the measurement.

While we can relatively easily identify physical objects, then, that we want to measure – the table, the building or the atom – this is not so clearly the case with language. Language, like intelligence, cognitive ability and personality, is a psychological construct, and its measurement can involve the assessment of many different aspects of language (e.g. broadly, the four macro skills of speaking, reading, writing and listening). In order to be able to begin to measure language knowledge, then, we need to know what language consists of; in other words, we need to have a model of language; we need to *define the language construct*.

Task A2.4



- Think about what it means to know a language. What do you need to know to be able to speak and use a language? Clearly, you need to know how the grammar works, how to pronounce the language, and the vocabulary of the language. But what else do you need to know?

Defining the language construct

What do we mean when we say that someone knows how to use a language (Spolsky 1985)? This is something to which language testers have given considerable thought because before measures can be developed, we must know what the construct is – in other words, we need to know what, exactly, it is that we are measuring.

Language is a complex phenomenon, and there are many different aspects which can be assessed. This can include the four skills – speaking, reading, writing and listening – and, more specifically, particular aspects of language – pronunciation, intelligibility, syntax, vocabulary and discourse skill, to name but a few. Another important aspect of testing that needs to be considered is that, when we are testing a person's language, we can only elicit samples of their language performance. It is never possible to do more than this, since we cannot hope to be able to sample anyone's language in all possible circumstances. In other words, in evaluating a person's linguistic performance, we have to make inferences about their ability to use the language in other contexts from those which we are able to sample in the test. It is, therefore, as Bachman and Palmer (1996: 66) argue, crucial to define what it is we mean by language ability:

If we are to make inferences about language ability on the basis of performance on language tests, we need to define this ability in sufficiently precise terms to distinguish it from other individual characteristics that can affect test performance. We also need to define language ability in a way that is appropriate for each particular testing situation, that is, for a specific purpose, group of test takers, and TLU [test language use] domain. For example, we may want to focus on test takers' knowledge of how to organise utterances to form texts for one particular testing situation, while

in another we may be more interested in their knowledge of appropriate politeness markers. The way we define language ability for a particular testing situation, then, becomes the basis for the kinds of inferences we can make from the test performance.



Task A2.5

- Think about the kind of language skills you need to have to complete an arts degree at university successfully. Make a list, in order of importance, of the language skills you believe would be required. How would the list differ if the degree were in engineering?

Probably the most widely cited model in language testing has been that of Bachman (1990), further developed in Bachman and Palmer (1996), who have proposed models of communicative language ability that take into account both *language knowledge* and *strategic competence*. *Language knowledge* includes the knowledge which is required to use language appropriately in particular contexts, and is made up of several components (Bachman and Palmer 1996). *Organizational knowledge* refers to the formal properties of language required for understanding and producing language, as well as organizing language into longer stretches. It consists of grammatical knowledge and textual knowledge:

Grammatical knowledge is involved in producing or comprehending formally accurate utterances or sentences. This includes knowledge of vocabulary, syntax, phonology, and graphology.

Textual knowledge is involved in producing and comprehending texts ... *spoken or written* – that consist of two or more utterances or sentences. There are two areas of textual knowledge: knowledge of cohesion and knowledge of rhetorical or conversational organization.

(Bachman and Palmer 1996: 68)

Pragmatic knowledge relates utterances, sentences or texts to the communicative goals of the people using the language and to the setting in which it is used. It can be divided into *functional knowledge* and *sociolinguistic knowledge*. Functional knowledge is made up of:

Knowledge of ideational functions [which] enables us to express or interpret meaning in terms of our experience of the real world. These functions include the use of language to express or exchange information about ideas, knowledge or feelings. Descriptions, classifications, explanations, and expressions of sorrow or anger are examples of utterances that perform ideational functions.

Knowledge of manipulative functions enables us to use language to affect the world around us. This includes knowledge of the following:

- 1 *instrumental functions*, which are performed to get other people to do things for us (examples include requests, suggestions, commands, and warnings)
- 2 *regulatory functions*, which are used to control what other people do (examples include rules, regulations, and laws); and
- 3 *interpersonal functions*, which are used to establish, maintain and change interpersonal relationships (examples include greetings and leave takings, compliments, insults, and apologies).

Knowledge of heuristic functions enables us to use language to extend our knowledge of the world around us, such as when we use language for teaching and learning, for problem-solving, and for the retention of information.

Knowledge of imaginative functions enables us to use language to create an imaginary world or extend the world around us for humorous or esthetic purposes; examples include jokes and the use of figurative language and poetry.

(Bachman and Palmer 1996: 69–70)

Finally there is *sociolinguistic knowledge*, which allows us to use language appropriate to the context in which it is being used – this may include knowledge of the most appropriate dialect to use in a particular situation, or when to use different registers, such as when to use a formal register versus when to use more informal registers.

This understanding of language in theoretical terms is critically important in testing because it is from this model of language that we develop our test construct – the abstract theoretical concept that is reflected in test performance. In order to test language, therefore, we need to (1) know how to *describe* language, and (2) know how to *operationalize* it in terms which are precise enough to allow it to be measured. This means we need to be able to break down language into measurable units as, generally, language competence is an invisible attribute. Obviously we cannot realistically measure every aspect of the language itself. Thus, we are inevitably going to *infer* language ability on the basis of what we are able to observe. In order to do this, we first need to identify what the abilities are that we want to measure.

In any assessment, there are two central concepts which are crucial to interpreting the outcome of the measurement. The first is *validity*. Validity refers to the extent to which the measurement instrument (e.g. the language test) is an appropriate measure of the phenomenon itself – in this case, language. For example, a vocabulary test could not be considered a valid measure of communicative ability in a language, but a rating of an oral interview would probably provide a more valid measure. The second is *reliability*. Reliability refers to the extent to which something is measured consistently. There are various ways of measuring reliability – for example, a test may be given twice to the same group of people to see the extent to which the scores are consistent across the two occasions. These issues are discussed in more depth in Fulcher and Davidson (2007).

Types of test

When we are measuring second-language proficiency (as we have defined elective bilingualism), there are various purposes for which tests can be used. Henning (1987) discusses a range of types of test, of which the following are probably the most commonly used:

- *Proficiency tests*: designed to measure a person's language ability irrespective of the type of language experiences the person may have had. These are tests of general language proficiency where the required language proficiency is specified by a set of expectations of the types of activities the candidate would be expected to be able to do with language in order to be considered a proficient speaker of that language. Two well-known proficiency tests are the IELTS (International English Language Testing System) and TOEFL (Test of English as a Foreign Language).
- *Achievement tests*: designed to evaluate the language learned in a specific language instruction programme; such tests relate specifically to the curriculum of the course and will be designed to evaluate either progress or the final achievement of the learners.
- *Diagnostic tests*: designed to identify areas of language strength and weakness, generally for the purposes of providing additional and appropriate assistance at a later time.
- *Placement tests*: designed to identify the most appropriate placement for learners in classes where classes of varying proficiency levels are available.

Both in language learning situations and in the testing and assessment of languages, we often think in terms of the four macro skills – speaking and writing (the productive skills) and reading and listening (the receptive skills) – and traditionally tests have often been designed to specifically evaluate each of these skills. Recently, however, there has been an increasing focus on the development of more integrated tasks which go beyond the notion of assessing each skill independently and require candidates to perform tasks in which more than one skill is tested (e.g. reading a text and then writing a summary of it tests both reading and writing skills) (Wigglesworth, in press).



Task A2.6

It is often the case that individuals do not demonstrate the same levels of competence in each of these four skills.

- Under what circumstances might individuals develop their skills in one area more than in another? See if you can think of some examples.
- Under what circumstances might you decide that:



- testing one or two skills is **more important** than to test another?
- testing one or two of them **might not** be necessary at all?

➤ Provide examples and give reasons for your choices.

Tests can vary on a number of different dimensions, and these dimensions have implications for both the validity and the reliability of the tests. While we cannot discuss these in any detail here, those interested in exploring the concepts underlying testing and assessment are directed to Bachman and Palmer (1996), Brown (2004), Henning (1987), Hughes (1989) and Fulcher and Davidson (2007). While all of these provide relatively in-depth treatments, a brief and accessible overview of testing concepts can be found in McNamara (2000).

Tests may be:

- direct or indirect
- scored objectively or subjectively
- criterion- or norm-referenced.

Direct tests test the actual skill under investigation, while *indirect tests* measure the abilities which underlie the skill in which we are interested (Hughes 1989). To illustrate this difference, Hughes uses the example of writing an essay. In a direct test, we would ask the candidate to write an essay, which would then be scored; in an indirect test, we might devise a set of grammar correction items in which the candidate was required to identify the incorrect item and revise it. While this would not be a direct test of the candidate's ability to write an essay, such skills are strongly correlated with the ability to write an essay.

Tests can be scored either *objectively* or *subjectively*. Objective scoring does not require any judgement on the part of the assessor. An example is a multiple-choice item in which the correct item has to be selected from a set of distractor items. In this case, the answer is either right (where the correct item is chosen) or wrong (where it is not). Subjective scoring refers to the type of assessment where judgements need to be made by a rater, which usually involves some measure of judgement and expertise. Written texts and spoken discourse are usually subjectively scored by a rater using a rating scale which defines different levels of proficiency. Rating scales may be either *holistic* – where the mark is assigned on the basis of a holistic evaluation of the candidate's speech or writing – or *analytic* – where several scores are assigned on the basis of different aspects of the speech or writing. For example, in an analytic rating scale for a speech sample, the rater might be asked to assign three scores, one for grammar, one for intelligibility (pronunciation and stress patterns) and one for breadth of vocabulary. Thus, a candidate's written essay or their response to an interviewer's questions would be examples of the type of activity which would be subjectively scored. *Reliability* is always of concern in subjective scoring since raters tend to vary in their harshness (Upshur and Turner 1999) and in the way they interpret the rating scale (Lumley 2002). Rater training

attempts to address problems with rater consistency and to ensure that raters assess as fairly as possible. In some cases, statistical analyses can be used to contribute to identifying rater bias (Wigglesworth 1993; McNamara 1996), but, because raters do tend to vary, best practice in testing would require each text to be scored by more than a single rater.

Any kind of measurement implies a comparison: this may be with other members of the group or cohort, where measurements are made and compared across the group, often in relation to a standardized set of results, or the comparison may be with some definition of ability, or criterion. The former is known as *norm referencing*, and the latter as *criterion referencing*.

Performance-based testing

A performance-based test is 'a test in which the ability of candidates to perform particular tasks, usually associated with job or study requirements, is assessed' (Davies et al. 1999: 114). These kinds of test are designed to measure a candidate's productive language skills through performances on tasks which allow the candidate to demonstrate the type of language skills which it is expected they will be required to use at some later stage. Thus, a central tenet of performance tests is authenticity, where tasks are designed to be authentic representations of the types of activities candidates will perform in the real world; however, the degree to which it is possible to make inferences based on the language elicited by particular test tasks as a reflection of the candidates' ability to manage the task in a subsequent real-world context remains a matter for debate (Wigglesworth, in press).

In many ways, performance tests appear to enhance validity by eliciting samples of the type of language that will be required in future situations, but Darling-Hammond (1994) offers a cautionary note about such tests, pointing out that performance-based assessments do not inherently mean that testing will be more equitable. She argues that we still need to apply rigorous standards and be very careful that such assessments really are fair and valid forms of assessment. These types of assessment may be seen as preferable because they are essentially seen as more authentic: they are based on the kind of real-world tasks that students may normally be expected to participate in, and evaluated on criteria which assess the actual performance, rather than the multiple-choice type, or discrete type test items which have traditionally been administered. However, as Darling-Hammond points out, we must still be vigilant with many of the traditional concerns with any kind of testing: for example, determining that they avoid bias, evaluating how they resolve concerns about subjectivity versus objectivity, and how they influence curriculum and teaching – that is, the washback effect that such assessments may have.

Alternative types of assessment

There are also alternative ways of making assessments about language knowledge, and these include portfolio assessments where individuals collect a number of samples of their work which demonstrate their language ability. As Diaz Rico and Weed (1995) suggest, such portfolio assessments may include written work, self-assessments, reports, dictations, audio recordings etc. and are designed to show the learner's progress over time. There are also competency-based assessments where language knowledge is evaluated in relation to the learner's ability to perform particular task competencies under various conditions. In these cases, although assessment may often be done formally, it is generally embedded within the curriculum. For example, in the Adult Migrant English programme in Australia, assessment is competency-based, and so learners may be required to complete a role-play activity in which they are asked to make a complex request (such as asking to change an interview time), or they might be asked to write a short report, for example on a memorable day in their lives.

Task A2.7



- Consider the following options for assessing the writing proficiency of a class of twelve-year-old second-language learners and then respond to the questions below:

Type 1 Portfolio assessment (consisting of three drafts of two essays of 500 words; two drafts of two argumentative essays; a journal; a weekly dictation; material collected over one term)

Type 2 An exam-based assessment consisting of two essays, one descriptive, one argumentative, of around 300 words, written under exam conditions

Type 3 A test of grammar, a cloze test and a spelling test conducted under exam conditions (a cloze test is a written test in which sentences or paragraphs are provided in which certain words have been deleted. The test candidate is required to fill in the spaces, e.g. 'Yesterday Mike ____ to the market and bought some carrots')

- Which type of assessment do you think:

- would be the most *valid* (in other words, the most accurate in predicting the proficiency level of the children's language ability)?
- would be the most *time-consuming*?
- would be the most *reliable*?
- would be the one the *children* preferred?
- would be the one the *teacher* preferred?

- Which would you choose in this situation? What are the advantages and disadvantages of each?

While more communicative models of language testing are currently prominent, there is also a move towards more holistic, qualitative assessment. These alternative modes of assessment offer more qualitative assessments of the learner's abilities on the whole, but also involve a more time-consuming commitment on the part of the assessor and the learner, and do not provide comparable data in the way that more traditional types of tests offer. They may also not be as equitable, and fairness and ethical concerns need to be taken seriously (McNamara 1996, 1998).

We have now briefly overviewed a range of considerations relevant to the ways in which language can be assessed formally. On the whole, language tests and language assessments are designed for elective bilinguals: that is, they evaluate the language of second-language learners against the norm of a native speaker of that language (Valdes and Figueroa 1994). Circumstantial bilinguals fall into a rather different category, and the question of whether they should be evaluated against native-speaker norms is one which has been widely discussed. Grosjean (1989) in a seminal article argued strongly for the view that a bilingual should not be seen as the sum of two monolinguals. Taking the point further, De Groot and Kroll (1997: 2) argue that bilingualism cannot be viewed as simply the sum of two monolingual minds but that we need to take account of the interaction between the two languages. Further, they argue that this interaction is a complex one, the investigation of which will require detailed understanding of monolingual knowledge of language as well as of the bilingual knowledge of the languages. This means that the issue of the relationship between the bilingual's two languages becomes one of paramount importance. As we will see below, these issues are very complex ones.

ASSESSING THE LANGUAGE OF CIRCUMSTANTIAL BILINGUALS

When we are assessing the language of elective bilinguals, we are not really concerned with their competence in their first language (or, in other words, the language of which they are a native speaker). When we turn to look at the situation with respect to circumstantial bilinguals, the situation – particularly with children – is rather different. We have chosen to focus on the assessment of children because of both the complexity and the importance of doing so – for example, 11 per cent of all primary, and 9.1 per cent of all secondary, schoolchildren in the UK come from non-English-speaking backgrounds (Strand and Demie 2005).

In Unit A1, we discussed extensively the concept of the 'domain' of the bilingual's two languages. This is an important distinction which often differentiates the elective bilingual from the circumstantial bilingual. As we have seen above, an elective bilingual is by nature a native speaker of a language. This means that generally

this person is a native speaker of a language and that this language will be spoken across all domains – at home, in the kitchen, with family, with relatives, at work, on the phone for giving lectures or presentations, in conferences, on the factory floor, in the community, at the shops, at social events, at the children's school etc.

When we are considering assessing circumstantial bilingual language, we need to think in terms of domains of language use. While these domains will not be linguistically differentiated for the monolingual speaker, it is frequently the case that they are differentiated for the circumstantial bilingual. With adult circumstantial bilinguals, we might want to consider whether they have been educated in one or both of their languages – and, related to that, what level of educational attainment has been achieved, and whether this was equivalent in both languages. It is, in fact, quite unusual for circumstantial bilinguals to use both languages in exactly equivalent domains, or for exactly equivalent functions. This point, therefore, needs to be taken into consideration. Many bilinguals, for example, may use only one of their languages in their professional or work life, while the other is used exclusively at home. In these cases, one of the languages may be dominant over the other.

Alternatively, it may be that one of the languages is more dominant in certain domains, while the other language is dominant in a different set of domains. Thus, a university lecturer may lecture in English on a topic (say geography) in an English-speaking country, but speak Japanese exclusively at home and in the community. This does not mean that if requested to give a lecture on geography in Japanese, this speaker will necessarily be able to do so with any ease. These language functions occur in different domains, and the skills required for one are not necessarily instantaneously transferable to the other (for example, particularly in a case like this where specialized vocabulary is required, the speaker may not have available this vocabulary in the 'home' language). This issue of the domain of language use is clearly one which needs to be taken into consideration since a circumstantial bilingual may not have equal competence in the performance of both their languages in all domains available. We saw in Unit A1 the importance of different domains of language use for bilinguals, so it is important that any measurement of the language use pertains to the domains in which those languages are used.

ASSESSING THE LANGUAGE OF CIRCUMSTANTIALLY BILINGUAL CHILDREN

In the last section of this unit, we are going to focus specifically on assessment issues as they pertain to children growing up in situations of circumstantial bilingualism. As Table A2.1 (p. 21) shows, children growing up as circumstantial bilinguals are generally children who come to school speaking a minority language of the country in which they are living. When they arrive in school, they may have had only limited access to the language of education in the community in which they are living. However, the importance of appropriate assessment of the needs of such children cannot be overestimated. This is because these children now make up a substantial

proportion of the school population. Mahon et al. (2003) point out that one in eight schoolchildren in the UK come from a minority ethnic background. In the US, the number of limited-English-proficient (LEP) students attending school increased by 65 per cent in the ten years from 1994 to 2004, with approximately one in ten schoolchildren now coming from such backgrounds. This is in striking contrast to the situation in many other countries, where there is enormous variation in language backgrounds, approximately 80 per cent of these LEP children report Spanish as their native language (Batalova 2006). Clearly, this makes the assessment of both languages a much more accessible goal.

The work on assessing the language of circumstantially bilingual children comes from two main sources. One is the assessment of such children in the educational system, where there is often quite substantial standardized testing of content material. The other is the assessment of children who may have specific language impairment, which should be diagnosed as early as possible so that they may be provided with appropriate treatment. We will deal with these issues briefly below. In Unit B2, we will consider in more depth those issues that we need to take into consideration when assessing bilingual children in the school situation.

The use of norm-referenced tests is widely recognized as being problematic for assessing children being raised in situations of circumstantial bilingualism (Genishi and Brainard 1995; Gutierrez-Clennen 1996; Thomas and Collier 1997), because such tests have generally been normed on populations which do not include bilingual children. For this reason, bilingual children who are tested with these instruments may be seriously disadvantaged and assessed as not only linguistically limited but also cognitively deficient (see Unit A4 for a more detailed discussion of how this has impacted on views of bilingualism and cognitive function).

The importance of appropriately assessing circumstantially bilingual children has been widely recognized because of the potential for language problems to result in poor school performance, and subsequently to limit life choices. Cummins (1979, 1984) makes a distinction between *basic interpersonal communicative skills* and *cognitive academic language proficiency* (more recently termed *conversational language proficiency* and *academic language proficiency* respectively in Cummins 2000). This view has been the subject of considerable discussion and controversy. Romaine (1995: 268) argues that the model is too simplistic, and empirically difficult to test, while Hamers and Blanc (2000: 99) maintain that it lacks explanatory adequacy since it does not address the issue of simultaneous bilingualism, nor does it explain why children fare differently. None the less, the model remains a useful one given the fact that as children go through the formal educational system, they need to acquire very different language skills from those that they are using routinely at home. The distinction between basic conversational proficiency and academic language proficiency is a very important one because, while children may reach basic conversational fluency within two years, it takes between five and seven years to attain academic language proficiency (Strand and Demie 2005). Thomas and Collier (1997) report on two earlier studies which suggest that these estimates may be rather conservative:

The student samples consisted of 1,548 and 2,014 immigrant students just beginning their acquisition of English, 65 percent of whom were of Asian descent and 20 percent of Hispanic descent, the rest representing 75 languages from around the world. These students received 1–3 hours per day of ESL instructional support, attending mainstream (grade-level) classes the remainder of the school day, and were generally exited from ESL within the first two years of their arrival in the U.S.

We limited our analyses to only those newly arriving immigrant students who were assessed when they arrived in this country as being at or above grade level in their home country schooling in native language, since we expected this ‘advantaged’ on-grade-level group to achieve academically in their second language in the shortest time possible. It was quite a surprise to find a similar 5–7 year pattern to that which Cummins found, for certain groups of students. We found that students who arrived between ages 8 and 11, who had received at least 2–5 years of schooling taught through their primary language (L1) in their home country, were the lucky ones who took only 5–7 years. Those who arrived before age 8 required 7–10 years or more! These children arriving during the early childhood years (before age 8) had the same background characteristics as the 8–11-year-old arrivals. The only difference between the two groups was that the younger children had received little or no formal schooling in their first language (L1), and this factor appeared to be a significant predictor in these first studies.

(Thomas and Collier 1997: 33)

Task A2.8



- Can you think of what types of activities children around the age of five would be involved in at home and make a list of them (e.g. playing, watching children’s programmes on television, shopping at the supermarket etc.)? What kind of language do you think would be required for these kinds of activities?
- When children begin to attend school, they must both behave in a different way and use language in different ways. List the kinds of activities children participate in at school which are different from those at home (e.g. learning a range of academic skills including numeracy and literacy, sitting at tables or on mats in groups for a reasonable period of time, listening to detailed instructions about activities). Can you describe the kind of language that would be required for these kinds of activities?

ASSESSING BILINGUAL CHILDREN FOR SPECIFIC LANGUAGE IMPAIRMENT

At any one time, 10 per cent of the population of normal children entering the school system will be affected by some sort of speech disorder; from this it follows that a similar proportion of bilingual children will be affected (Holm et al. 1999). The likelihood is that when children or their parents are advised that the child should see a speech pathologist they may encounter a number of difficulties. These include the high chance that the speech pathologist will not be familiar with at least one of the child's languages, and standardized (or any other kind of) assessment instruments will not be available in at least one of their languages. Where assessment instruments are available, they may not be in the child's dominant language, and there may be very limited norms available for bilingual children.

This potential for bias in language assessment for linguistic and cultural minority students has been widely recognized (Valdes and Figueroa 1994). Saenz and Huer (2003) point out that second-language students routinely score lower than their monolingual peers despite the appearance of fluency at the conversational level. While the standardization procedure may include some children from minorities, they are none the less biased, usually towards the middle-class monolingual English children who constitute the majority of the population upon whom such tests are standardized. Furthermore, Pert and Letts (2003) argue strongly for the importance of ensuring that when children are being tested for language impairments, all the languages the child uses are tested. This is because using English-only procedures will not be adequate to assess the child's full linguistic repertoire, and this is particularly the case where the child has had only limited opportunities to learn English. Pert and Letts describe the kinds of problems which are encountered:

Even if a therapist is bilingual, they are unlikely to speak all the languages they encounter (Lahey, 1992), and the lack of normative data means that diagnosis will be based on subjective decision making. Clinicians have been severely hampered by the lack of assessment materials available in languages other than English, and often by lack of access to appropriate language speakers who possess the skills to administer an assessment. The latter problem has been overcome in some areas by the employment and training of bilingual co-workers, without whom an effective service for non-English first language students would be impossible to sustain. Co-workers can also provide information about the languages they speak, which will help the therapist make judgements about the normality or otherwise of any particular child's language skills. In the absence of test materials, the clinician can use spontaneous speech sampling, aided by a co-worker, to form an impression of the child's expressive skills when speaking his or her first language. However, in contrast to English, there is no guarantee that the therapist will have access to developmental stages and norms for the target language. Without a body of research on which to base decisions about developmental stages, the therapist who is

professionally committed to providing an accurate assessment and to planning intervention is effectively attempting to do the impossible.

(2003: 268)

The study undertaken by Pert and Letts tries to address these issues, and they raise a number of points which must be taken into consideration when testing children from bilingual backgrounds, as well as highlighting the complexities involved in this kind of testing. In discussing the development of tests of expressive language in three different Asian languages widely used in their region of the UK, they point out a few of the difficulties which may be encountered in attempting to address these issues. The first is that if the home, or first, language is not considered to carry much prestige with it, then parents may not report it as a language they use. They also point to the problem of identifying different dialects which may have very similar names. Finally, Martin et al. (2003) suggest that the 'standard variety' of the language spoken in the original home country may not be the same as that spoken by second- and third-generation migrants now living elsewhere.

Assessments used with children also need to be age-appropriate, and should not be confronting for the child. Hasselgreen (2005) points out that there is now general consensus on the features that assessment activities for young children should encapsulate. These are:

- Tasks should be appealing to the age group, interesting and captivating, preferably with elements of game and fun
- many types of assessment should be used, with the pupil's, the parents' and the teachers' perspectives involved
- both the tasks and the forms of feedback should be designed so that the pupil's strengths (what he or she can do) are highlighted
- the pupil should, at least under some circumstances, be given support in carrying out the tasks
- the teacher should be given access to and support in understanding basic criteria and methods for assessing language ability
- the activities used in an assessment should be good learning activities in themselves.

(Hasselgreen 2005: 338–339)

SELF-ASSESSMENTS OF BILINGUAL PROFICIENCY

There are two ways in which we can obtain information about a bilingual's language from self-assessments. One approach is through the use of language background scales, such as the one given in Task C1.4 (pp. 271–273) in which bilinguals are asked to provide information about their language use with a series of questions which investigate to whom they speak each of their languages, and how often they do so. Baker (2006: 33–34) points out that these kinds of scales necessarily have some limitations because they are generally

not exhaustive of targets (people) or of domains (contexts). Language activity with uncles and aunts, discos, correspondence, organizations, hobbies and travel are not included, for example. The choice of items included in such a scale is somewhere between an all-inclusive scale and a more narrow sample of major domains. At first glance, it may appear the more inclusive a scale is the better. There is a problem, illustrated by Baker & Hinde (1984: 46): 'A person who says she speaks Welsh to her father (mostly away at sea), her grandparents (seen once a year), her friends (but tends to be an isolate), reads Welsh books and newspapers (only occasionally), attends Welsh Chapel (marriages and funerals only) but spends most of her time with an English speaking mother and in an English speaking school might gain a fairly high "Welsh" score'.



Task A2.9

- Refer to the language background questionnaire provided in Task C1.4. How well would this work with the person described in the example above? How might these limitations be addressed without ending up with an exhaustive, but massively long, questionnaire? To do this it will help to think about ways in which you might prioritize questions about particular people or activities.

While language background scales of the type discussed above have been used extensively in bilingual research, they do not purport to measure the proficiency of speakers, rather their patterns of language use across a variety of domains. Proficiency can, however, be measured through self-assessment questionnaires, and although these have limitations they can serve a useful purpose. Little (2005: 321–322) argues that there are three main reasons for using self-assessments: firstly, for learners to be able to assess their progress in terms of the curriculum they are learning; secondly, to encourage learners to regard assessment as a shared responsibility; and thirdly, to allow learners to identify occasions in which the target language can be used for additional explicit language learning. While self-assessment of language skill is generally related to curriculum concerns, and closely tied to it (see Ross 1998 for a detailed discussion of this), the Council of Europe has now developed a self-assessment instrument called the European Language Passport as part of the European Language Portfolio. This allows speakers to evaluate their language skills in terms of understanding (reading and listening), speaking (spoken interaction and spoken production) and writing. Skills in each of the five areas are assessed on a six-point scale (A1, A2, B1, B2, C1, C2) where A is advanced, B intermediate and C beginner. At the end of the process, the 'Language Passport' can be downloaded (see <http://europass.cedefop.europa.eu>). 'Dialang' (www.dialang.org) is another site where you can test your proficiencies in fourteen different languages.

Summary

In this introduction to the complexities of measuring bilingualism, we began with an overview of the various types of bilinguals we might want to assess. We focused on the evaluation of the language firstly of elective bilinguals, where our interest will be on assessing them in their second language. We then looked at the more complex question of assessing the language of circumstantially bilingual individuals, specifically considering the difficulty of assessing children in these situations.

In the Extension part (Unit B2), we will build on the discussion in this unit by reading three extracts; the first, by Saenz and Huer (2003), posits a variety of alternative forms of testing, showing the advantages and disadvantages of each, focused particularly on speech pathologists. The second reading, by Bialystok (2001a), discusses the issue of proficiency and what it means in different circumstances. The third focuses on a widely cited theoretical model proposed by Cummins (2000) designed to raise our awareness of the kinds of factors which need to be taken into account in assessing children's language in educational systems.