

## Unit B2

### Measuring bilingualism extended

In Unit A2, we considered some of the key concepts which need to be taken into consideration when we assess languages. We discussed the importance of identifying different groups of bilinguals and considered also the ways in which we could assess an individual's developing bilingualism. We looked at this differently depending on which of two different types of bilinguals we were concerned with assessing: elective or circumstantial bilinguals. In this Extension section, we are going to consider in more depth some of the concerns that we have been discussing in relation to assessing the language of bilingual children in a range of circumstances.

#### ASSESSING THE LANGUAGE OF BILINGUAL CHILDREN

We will focus on children because the assessment of children's language is required in many contexts. One reason for assessing a child's language can be to identify whether there are any inherent problems in the child's language development which may require attention for the purposes of remediation. We will begin with a reading in which this issue is considered further. Children's language also often needs to be assessed for educational purposes. In the second article, we consider how language proficiency may impact on children both educationally and cognitively, and in the third reading we examine the role of language proficiency in the classroom and the ways in which it may be assessed.

In Unit A2, we discussed some of the difficulties with developing appropriate assessments for children who may have developmental delays in both of their languages. In the following extract, Saenz and Huer (2003) discuss some of the practical approaches that can be taken to alleviate this problem. We saw that there was a serious potential for bias in the testing of bilingual children, and this first extract considers the advantages and disadvantages of a variety of alternatives to standardized tests, with a view to moving towards less biased testing practices for children who speak minority languages. While the extract focuses on children with potential developmental speech delay, many of the points raised are as relevant to more educationally oriented testing as they are to the identification of delays or difficulties in language development.

Saenz and Huer (2003) discuss in some detail the kinds of approaches that can be taken to ensure that bilingual children are having their language assessed by the least



biased means possible. They do this by adopting various methodologies which can be used in the assessment process for bilingual children. The approach selected will depend on a range of factors and the practicalities of adopting one approach over another. *Renorming*, for example, in which a standardized test is administered to a different, and more appropriate, group from that on which it was originally standardized, may be a relatively expensive option since it involves considerable administration and analysis costs, whereas dynamic testing may well prove a cheaper option but requires a considerable time commitment for each individual tested.

Text B2.1  
T.I. Saenz and  
M.B. Huer

**T.I. Saenz and M.B. Huer (2003) Testing strategies involving least biased language assessment of bilingual children. *Communication Disorders Quarterly*, 24 (4): 184–193.**

### Renorming

Renorming a standardized test using the target population at issue has been advocated as an alternative to using norms based upon the population of the United States. The issue of establishing appropriate norms is an important one. Genesee and Upshur (1996) stated that an appropriate norming group should include individuals of the same ethnic and cultural backgrounds, linguistic background, age, gender, and educational level and type. The group on which the test is normed should also be relatively homogeneous. . . .

Renorming a test has several advantages. In a large and relatively homogeneous school district like Santa Ana, large numbers of bilingual pupils share the same cultural and linguistic backgrounds and economic status, making the time and effort expended in renorming potentially worthwhile. When the performance of particular students in the district is compared to that of their peers, speech-language clinicians can be relatively confident that students who score in the low range on the test rank substantially below peers of the same background.

Renorming also possesses a number of disadvantages. The process requires substantial coordination and time to assess enough typically developing children to provide meaningful norms. In a small or relatively unpopulated geographical area or highly diverse school district, there may not be enough speech-language pathologists or students sharing the same cultural/linguistic backgrounds to collect meaningful local norms. In addition, applicability of the norms is limited to the population upon which it has been renormed (Blevins, personal communication, November 7, 2002).

Tests that are both translated and renormed also should have their reliability and validity established for the intended uses (Baker et al., 1998). The difficulty in finding exact equivalencies for English words and concepts may change the absolute and relative difficulty of test items, seriously compromising a test's reliability (Kayser, 1998; Roseberry-McKibbin, 2002). Translated items also may not correspond as clearly with pictured items on a vocabulary test as the original English vocabulary words. In addition, grammatical structures in different languages may not be equivalent or may be acquired at different ages. Finally, differences in life experiences may not easily translate from one culture to another (Roseberry-McKibbin, 2002).

Another concern about local norms is that they frequently are lower than the norms for the samples tested during a test's initial standardization (Brice, 2002). Although

such norms may be considered appropriate for local use, individuals rightly may be concerned that separate publication of norms for different cultural or linguistic groups may (a) give the impression that some groups are not as linguistically capable as others and (b) be used to question the competency of students from different backgrounds.

One factor in lower scores is the familiarity of the testing task to the children who are being tested and the way in which test items are presented. Peña, Quinn, and Iglesias (1992) discovered that low-income Puerto Rican children in the United States often were not taught labels of objects, making labeling an unfamiliar task. Peña and Quinn (1997) also found that Puerto Rican and African American children performed significantly better on the Comprehension subtest of the *Stanford-Binet Intelligence Test* (Thorndike, Hagen, & Sattler, 1986), which required descriptions, than on the *Expressive One-Word Picture Vocabulary Test* (EOWPVT; Gardner, 1979), which required object labeling. The more familiar description task was better than the labeling task at differentiating between children with typical skills and children with language disorders.

Consequently, lower performance on renormed language tests by children who are culturally or linguistically diverse may be due to a variety of factors, including family income, difficulties with translation, familiarity and presentation of tasks, and fluency in English. A test that attempts to provide differential diagnosis thus may have a number of confounding factors that affect a child's score. Although children should not be identified as having a disability based on linguistic, cultural, or economic factors, the influence of each may be difficult to calculate.

### Task B2.1.1

- Recall from Unit A2 that the immigrant population in the United States (which is where Santa Ana is) is much more homogeneous than the immigrant populations of many other countries because approximately 80 per cent of the children from a minority background are Spanish speakers. Why might renorming a test for a homogeneous population be more appropriate than for a heterogeneous one?

### Dynamic assessment

A second proposed solution to the issue of least biased assessment is dynamic assessment (Bain & Olswang, 1995; Brice, 2002; Goldstein, 2000; Jacobs, 2001; Kayser, 1998; Lidz & Peña, 1996; Peña & Quinn, 1997; Peña et al., 2001; Peña et al., 1992; Tzuriel, 2001). This method typically involves a sequence of testing, including pretesting, teaching, and posttesting. A child is first pretested using a standardized test or informal task. Following administration of the test, the evaluator works with the child as a coach or mentor on tasks similar to those on the pretest, teaching the child test strategies and techniques and working with the child on problems and concepts. The aim of this intervention phase is to teach the child techniques or concepts that will enable him or her to perform well on the posttest. Such components as the amount of examiner effort needed to help the student, the rapidity with which the child changes responses following teaching, and the generalization of task performance to other tasks can be measured (Goldstein, 2000). Following this phase, the child is

Text B2.1

T.L. Saenz and  
M.B. Huer

Text B2.1

T.L. Saenz and  
M.B. Huer



Text B2.1

T.I. Saenz and  
M.B. Huer

retested, using the same test as the pretest. The child's scores on the pretest and posttest can also be compared to determine the child's degree of transfer. According to the theory of dynamic assessment, children with typical language ability and poor pretest results may score substantially higher on the posttest because they have a typical ability to learn the test's concepts or procedures when given adequate exposure and effective teaching (Peña et al., 1992).

Dynamic assessment has a number of advantages (Peña et al., 2001; Peña et al., 1992; Ukrainetz et al., 2000). For example, after implementing a program introducing Puerto Rican preschoolers to labels, Peña et al. (2001) found that the mean scores of nondisabled Puerto Rican preschoolers on the EOWPVT rose from 1 standard deviation below the mean to within normal limits. Children with speech or language disorders, however, made far more modest gains, confirming the value of dynamic assessment techniques to differential diagnosis. Similarly, in a study of Native American preschoolers, Ukrainetz and colleagues discovered that children's categorization skills markedly improved after an intervention phase. In addition, because test administration is not modified, standardized test scores can be reported, and the students' rise in test scores, as well as their test scores following teaching, may be considered as important as initial test results (Peña et al., 2001).

Dynamic assessment does have some disadvantages, the greatest of which may be time (Brice, 2002; Peña, 1996). In most cases, using dynamic assessment involves pretest, posttest, and teaching phases, rather than one test administration. If the evaluator chooses a high-structure, standardized approach to dynamic assessment, he or she may need to develop intervention activities and cues for the intervention phase that are designed to teach the tasks measured by each test (Peña, 1996). Furthermore, the evaluator can use triple or more the time expended for one administration of the test in performing dynamic assessment, although innovations such as incorporating teaching procedures and graduated prompts into administration may reduce the time required (Bain & Olswang, 1995; Jacobs, 2001).

The amount of time and effort spent in teaching tasks in the intervention phase may be hard to justify for some tests. For example, both Peña et al. (1992) and Ukrainetz and colleagues (2000) selected skills important to academic achievement – labeling and categorization. Both skills had implications for children's success in school settings and thus would interface well with the current emphasis upon teaching skills relevant to classroom success. In contrast, the frequently used *Clinical Effectiveness of Language Functioning–Third Edition* (CELF-3; Semel, Wiig, & Secord, 1995) has six subtests for every age group tested, some of which test skills seldom directly used or taught in speech-language interventions. Because the most important score on the test is based on a combination of all six subtest scores, an individual attempting to perform dynamic assessment might need to provide interventions for each subtest. Consequently, dynamic assessment may be an effective measure of students' speech and language disorders, but time constraints may make it difficult to implement dynamic procedures with many tests or many students.

Other issues with dynamic assessment include reliability and validity. In low-structure, prescriptive approaches to dynamic assessment, cueing may not be standardized; therefore, increases in scores from pretest to posttest may not be comparable to those of other students (Peña, 1996). In such cases, the results of dynamic assessment may be better used to prescribe interventions than to determine if a child has a speech or language disorder.



### Task B2.1.2



- Dynamic assessment, because it involves a teaching phase, is a rather time-consuming approach to assessment. However, its great strength lies in its ability to assess the potential for learning of individual children. Given this, see if you can think of a couple of circumstances under which it might be a useful approach.

### Other nonstandardized measures

Text B2.1  
T.J. Saenz and  
M.B. Huettner

A third proposed alternative to standardized testing involves nonstandardized measures other than dynamic assessment, and researchers have proposed a variety of these. For example, parent interviews, questionnaires, and/or checklists (Cheng, 1991; Roseberry-McKibbin, 2002) may be used to obtain information about parental perspectives regarding a child's language disorder. Restrepo (1998) also found that parental report of a child's speech and language problems, number of errors per T unit,<sup>1</sup> family history of speech and language problems, and mean length of T unit are useful in discriminating between Spanish-speaking children with language deficits and Spanish-speaking children with typical language skills. The first two factors are especially useful in identifying children with language deficits.

Ethnographic assessment (Brice, 2002; Goldstein, 2000) involves the observation of a child in different contexts with different communication partners and is used to study the child's language abilities in naturalistic situations. Additional information may be obtained from interviews with parents, teachers, and aides in which the interviewee is encouraged to describe his or her experiences with and opinions about the child's speech and language skills (Westby, Burda, & Mehta, 2003). *Transdisciplinary Play-Based Assessment* (Linder, 1993), used with children up to age 7 years, includes some aspects of ethnographic assessment in a team approach. One member of the assessment team facilitates play with the child and elicits different behaviors while another member interviews the parent, using questions similar to those employed in ethnographic interviewing. A third team member videotapes the interaction, and all of the additional team members observe the facilitator-child interaction while using checklists to identify relevant behaviors.

A number of tasks related to classroom competencies have also been suggested as possible assessment alternatives. Gutierrez-Clennen (1998) found narratives and complex syntax and formulation tasks to be useful in eliciting the type of language common to classroom discourse. Criterion-referenced testing (Goldstein, 2000) has been used to analyze a child's mastery of key language behaviors instead of comparing the child's performance to that of other children. The key language behaviors selected for analysis may include tasks used in the curriculum. Portfolio assessment (Goldstein, 2000; Kayser, 1998) involves the collection over a period of time of a body of the child's work that represents different tasks and assignments. Portfolios may include samples of reading and writing abilities. Curriculum-based assessment (Baker et al., 1998; Shinn, Collins, & Gallagher, 1998) involves the use of tasks that

<sup>1</sup> A T-unit is a standardized measure of language defined by Hunt (1966: 735) as 'one main clause plus whatever subordinate clauses happen to be attached to or embedded within it'. [Editors' note]

Text B2.1  
T.I. Saenz and  
M.B. Huer

are similar to those used in the classroom to assess the skills necessary for success in school.

Informal measures have a variety of advantages. They frequently involve tasks that are more naturalistic than those performed in standardized tests and, consequently, are more directly related to success in the classroom. Because they are non-standardized, informal measures may be modified without substantially changing their validity, and they also can be designed and administered to effectively assess individual students' areas of strength or weakness.

Informal measures also have a variety of potential drawbacks. Because there are no norms, a student's performance cannot be quantitatively compared with that of other students. Reliability also may be an issue. Although research may establish some expectations for the acquisition of speech-language skills at specific ages, these expectations are typically based upon the performance of monolingual children. Consequently, the speech-language pathologist must rely on clinical experience to make a differential diagnosis. A speech-language pathologist who has extensive experience with children from a specific socioeconomic, linguistic, and cultural group may be able to make seasoned and appropriate clinical judgments; however, a less experienced speech-language pathologist – or one with little experience with a specific linguistic or cultural group – may have difficulty comparing a student's performance to the performances of other children from the same background (Langdon & Saenz, 1996). In addition, examiner bias may affect students' performance. Problems in this area include an overinterpretation of the student's abilities based on a small sampling of behavior, a lack of familiarity with linguistic and cultural issues that may affect the student's performance, and low expectations that result in not providing the student with sufficient opportunities to respond (Roseberry-McKibbin, 2002).



### Task B2.1.3

In the preceding paragraph, the following points are identified as potentially problematic in informal testing of children:

- reliability
- comparing bilingual children with monolingual children
- differential experience of speech pathologists
- lack of familiarity with cultural and linguistic issues affecting student performance
- low expectations.

➤ Can you think of any ways in which some of these concerns might be mitigated?

Text B2.1  
T.I. Saenz and  
M.B. Huer

### Test modification

The final alternative to standardized testing is test modification (Brice, 2002; Erickson & Iglesias, 1986; Goldstein, 2000; Kayser, 1998; Langdon & Saenz, 1996; Mattes & Omark, 1991; Roseberry-McKibbin, 2002). Based in part upon the idea that some bilingual students' lower performance on standardized tests is due to a lack of familiarity with testing tasks, test modifications are designed to provide further

instruction in and information about the task to enable the student to perform more effectively. Erickson and Iglesias, Mattes and Omark, Langdon and Saenz, Kayser, and Brice have advocated a variety of modifications of testing. Langdon and Saenz adapted Mattes and Omark's suggestions and made the following recommendations:

1. Make sure that students have had experience with content or tasks assessed by the test.
2. Reword or expand instructions.
3. Provide additional time to respond.
4. Test beyond the ceiling.
5. Record all responses and prompts.
6. Provide credit for responses that use a different dialect or language from Standard American English.
7. Use additional demonstration items.
8. Allow students to label items in receptive vocabulary tests to determine appropriateness of stimuli.
9. Have students explain incorrect answers.
10. Ask students to identify actual objects or items if they have had limited experience with books and pictures.

Although norms cannot be used in modified testing, proponents believe that test modifications can act as an abbreviated form of intervention and enable students who perform poorly on a standardized administration of the test to perform better on a modified version.

One of the main advantages of test modification is its time efficiency. Evaluators can choose to administer a test using standardized, and then modified, administration procedures. Alternatively, they may opt to use only modified administration procedures if norms are not to be used. One of the disadvantages is that unlike dynamic assessment, the second administration of a test cannot be scored using norms, because of the modification of testing procedures. Only the discrepancy between first and second test administration may be considered. In addition, many of the modifications also may need to be developed prior to test administration, necessitating an additional expenditure of time. Yet another disadvantage is that the efficacy of using test modifications has not been formally validated.

As you have seen from the extract, there are a variety of alternatives to using standardized and normed tests which address the issue of bias in the testing of children from linguistically and culturally diverse backgrounds. As is the case with any form of testing, however, nothing is perfect and each alternative has advantages and disadvantages. Which test to be adopted will depend on the purpose of the test, the time available for testing, and which alternative the clinician or educator considers is the most appropriate under the circumstances. As you can imagine, the assessment of children from linguistically and culturally diverse backgrounds is a sensitive issue because testing, by its nature, provides a measure of performance at a specific time and place, and may carry with it an element of error.

Text B2.1  
T.L. Saenz and  
M.B. Huer



### Task B2.1.4

- Imagine you had a friend who was raising their three-year-old child bilingually and who was worried about their child's language development. What kind of advice would you give your friend to help them to find an appropriate place to have their child's language assessed?

We began with this extract from Saenz and Huer (2003) for two reasons. It provides a useful practical checklist of the range of factors which need to be taken into consideration when testing bilingual children in both clinical and educational settings. Bearing in mind the concerns raised above, from this point we are going to move beyond such practical concerns about the assessment of children's language to a focus on some more theoretical considerations. In the next two extracts, we will examine some of the ways in which different levels of proficiency might impact on children both in terms of their cognitive development and in educational settings.

## THE ROLE OF BILINGUAL PROFICIENCY

The next extract, from Bialystok (2001a), considers the ways in which different levels of language proficiency may impact on educational attainment. This extract not only raises and discusses some important conceptual considerations in relation to bilingual proficiency but will also provide a background to the concepts and considerations encountered in Unit A4 on cognitive development. Bialystok's interest is in the cognitive and linguistic development of bilingual children, and in the extract below she discusses the role of proficiency, building on Cummins's work and the relationship of his model to academic success. The Threshold hypothesis (Cummins 1979) postulated that there are minimum levels of competency which have to be attained before the benefits of bilingualism can set in. He proposed two thresholds of language competence and argued that, to avoid negative effects from bilingualism, the lower threshold must be attained. Children at the first threshold would not experience negative or positive benefits from bilingualism. The cognitive growth of children who failed to reach the first threshold would be adversely affected. Those who attained the second threshold (high levels of bilingual competence) would enjoy all the enhancing effects of bilingualism.

Text B2.2  
E. Bialystok

**E. Bialystok (2001) *Bilingualism and development: Language, literacy, and cognition*. Cambridge: Cambridge University Press, pp. 226–231.**

There are two dimensions on which proficiency is pertinent to the development of bilingual children. The first is the *absolute* level of proficiency in each language, that is, a judgment of the child's competence relative to native speakers; the second is the *relative* level of proficiency across the two languages, that is, an estimate of the command of each of the bilingual's languages with respect to the other. Both of these have been investigated, although the distinction between them has not always been made clear.

The threshold hypothesis proposed by Cummins (1979) is a formal attempt to incorporate proficiency levels into predictions about the effects of bilingualism. The hypothesis is primarily driven by absolute levels, but it is framed within a context that places some significance on relative proficiency. The dependent variable of the hypothesis is academic success. Three types of bilingualism, based on both the relative and absolute mastery of the two languages, are identified. In limited bilingualism, the child lacks age-appropriate skills in both languages. In partial bilingualism, the child has achieved age-appropriate proficiency in one of the two languages. Finally, proficient bilingualism is marked by normal levels in both languages. He argues that limited bilingualism leads to cognitive and academic deficits but that proficient bilingualism results in cognitive advantages.

Cummins claims that no consequences follow from partial bilingualism, but it is on this point that absolute and relative proficiency levels make different predictions. For the measures that Cummins examines, primarily academic achievement, the most critical element is that children have age-appropriate levels in one of their languages and that this language gives them access to the abstract concepts that are the currency of schooling. Therefore, the relevant contrast is between children who have such facility in at least one language (partial bilingualism and proficient bilingualism) and children who do not (limited bilingualism). In contrast, research into cognitive, but not necessarily academic, consequences of bilingualism has used more subtle and arguably more arcane instruments of assessment. In this research, the important question is whether bilingualism adds benefits beyond those enjoyed by comparable children, not those at risk. The relevant contrast groups here are bilingual children with age-appropriate language skills in one language (partial bilingualism) versus bilingual children with age-appropriate language skills in both languages (proficient bilingualism). None of these children is at risk of academic failure, as Cummins rightly points out. Nonetheless, their bilingualism has had different impacts on subtle aspects of their developing cognitive profiles.

This is an important point as Bialystok here makes a very useful distinction which we can relate to both the different circumstances and different environments in which bilingual children are raised. Cummins's concern (children with limited bilingualism versus children with more proficient bilingualism) is with children in the school situation, the majority of whom will come from circumstantial situations, and for whom limited bilingualism can be a serious risk. Bialystok's concern is more research-oriented, focusing on the extent to which bilingualism in children can impact on cognitive development, and in this situation the focus is on partial versus proficient bilingualism. These concerns are taken up and explored in greater detail in Unit A4.

Some research with bilingual children who are not balanced in their relative language proficiency but are sufficiently skilled in one of the languages (cf. partial bilinguals) has sometimes reported positive cognitive effects for these children (e.g., Bialystok, 1988; Cromdal, 1999; Hakuta & Diaz, 1985; Yelland, Pollard, & Mercuri, 1993). In fact, Hakuta and Diaz (1985) showed that the greatest gains to cognitive and metalinguistic insights came in the earliest stages of being bilingual, when proficiency in the two languages was most asymmetrical. None of this research contradicts the point made by Cummins regarding the essential need for language proficiency to minimally place children in the domain of having access to academic language skills. The important

Text B2.2

E. Bialystok

Text B2.2

E. Bialystok

Text B2.2  
E. Bialystok

point is that children have not been found to suffer any disadvantages from learning and using two languages, even in academic settings, providing that one of their languages is established to a level appropriate for children their age. McLaughlin (1978) even claims that there may be no negative consequence if the instruction is carried out in the child's weaker language, the one that is not developed to age-appropriate levels (contradicting the warnings of Macnamara, 1966), as long as the level is *sufficient* to function in the instructional setting.



### Task B2.2.1

- Consider the point made by McLaughlin (1978) that there may be no negative consequences for the child if the instruction is in the weaker language '*as long as the level is sufficient to function in the instructional setting*'. What do you think are the implications of this statement for children arriving in their new country at different ages?

Text B2.2  
E. Bialystok

A related question is how skills developed in one language can be transferred to the other. The question is particularly important for children who have one stronger and one weaker language, and even more crucial when schooling is in the weaker language. Again, it is Cummins (1991) who has contributed significantly to this question. He reviews a large number of studies that compare the oral and academic language proficiency of children in their first language and the level of skill they acquire in each of these domains in their second language. These studies cover many language pairs and many instructional models. For academic uses of language, such as reading, there tend to be moderately strong correlations between the level of skill attained in both languages. Cummins attributes these relationships to the underlying cognitive involvement in development of these skills. For oral uses of language, the correspondence between children's competence in the two languages depends more on individual child attributes, such as personality and interaction style.

Part of the explanation for the transfer of skills across languages, as well as their lack of transfer, can be found again in the threshold hypothesis. Cummins' (1979) review of the literature shows that the detrimental effect of the lower threshold (lack of adequate competence in both languages) on academic achievement is well established. The reason is that there is no language in which children are able to establish the cognitive systems that are at the base of academic functioning. Even setting up these concepts in one language would permit their transfer to the other. The principle is the same whether the child is learning one or two languages: cognitive structures require the establishment of particular concepts, and these concepts require linguistic support.

These relations may appear circular but they illustrate that both absolute and relative proficiency levels are decisive in determining how the bilingual experience will affect children's academic and cognitive development. The positive cognitive consequences of bilingualism emerge when children control a reasonably balanced and competent mastery of the two languages. Although some cognitive processes have been shown to be enhanced from asymmetrical configurations of language competence, these effects are less reliable than are those that accompany regular and equivalent use of two languages. These effects of balance between languages appear more decisive than does the effect of time, namely, whether the two languages were

acquired together. Admittedly, however, no research has established this. As long as children ultimately master the two languages across the relevant contexts of use (home, school), it should not matter that they learned one before the other. Even under ideal conditions, bilingual children go through different periods of favoring one or the other of their languages. Therefore, to some extent the notion of simultaneity in the acquisition of two languages is irrelevant.

Text B2.2  
E. Bialystok

The point that Bialystok raises here is in relation to whether children learn both their languages simultaneously (i.e. from birth) or not. In the next unit, we will be looking in some detail at the acquisition of language bilingually and we will be focusing on children learning their languages simultaneously. However, as Bialystok points out, often we are not concerned with how children have acquired their bilingual state; 'a child of six or seven years old who becomes bilingual is a bilingual child' (Bialystok 2001a: 226). It does not matter whether children have learned two languages from birth, or whether they have learned one language and then the other sequentially.

For education to either profit from children's bilingualism or escape impairment because of it, the absolute proficiency level of at least one language is crucial. Considering only educational outcomes, the balance between the two languages is less critical than is the need for one of them to be developed to a level that is sufficient for schooling. But again, the timing for achieving that proficiency does not seem to matter.

Text B2.2  
E. Bialystok

Timing and proficiency, therefore, are two dimensions that distinguish among some of the configurations in which children learn two languages. Timing, the question of whether the languages should be learned one at a time or together, is probably not very important in the long run. Proficiency, however, does matter. Therefore, we can return to the question of the likely impact on a child who is moved to a different country and begins school in a different language from that spoken in the home. The child's weaker language may be either the one spoken at home or the one used in school. It is improbable that any academic consequence would follow from having some, even minimal, competence in a language that was not the language of schooling, but it is very consequential to consider the effect of having limited competence in the language of school.

Bialystok (2001a: 229) argues that in situations where children speak two languages at home and are educated in one of those languages (which is well established), there are never disadvantages for bilingualism, although there may only be advantages in specific situations. However, as we have seen, this is not always the case. In the next section she considers children in bilingual education programmes.

Consider the following situation: An educated middle-class family living in Berlin is about to move to Los Angeles because of a career opportunity for the mother. (Let's assume she was offered an important chair at a famous university and her husband has agreed to find work as a sales clerk or, better still, to stay at home and take care of the house). Their only child is four years old and speaks only German. The parents will continue to speak German at home because that is what is natural, but the child will enter school and begin to learn English. What will happen to the child and what should the parents do?

Text B2.2  
E. Bialystok

Text B2.2  
E. Bialystok

The indications are that the child will do just fine. She will learn English from the environment (because she will likely socialize with English-speaking children), from television, and from kindergarten. Kindergarten will require sufficient English to build the academic foundation for her education. By the time she starts first grade, she will probably have enough English upon which to build an academic base, although her English skills will be weaker than those of her native-speaking classmates. While her classmates are learning about letters, numbers, and the names of dinosaurs, she will additionally be learning the forms and structures of English, but the burden will not be onerous. If it is, or if she is slightly older and has not had the time to master English to levels appropriate for the grade, then some ESL training would be helpful. The parents will, of course, be worried that the child's English places her at a disadvantage and possibly consider imposing English as the language of the home.



### Task B2.2.2

- As Bialystok goes on to say, given the environment in which the child lives, and with the educational circumstances she is in, there is no reason to believe that the child's English will not develop 'in due course' or that it will in any way detract from her being academically successful. If you were talking to her parents, what reasons would you give to support this view?

Text B2.2  
E. Bialystok

### The learning latency

But what does 'in due course' mean? How long should it take children to acquire a level of proficiency in the school language that allows them to function and thrive in an academic environment? Many of our attitudes, expectations, and policies are based on some implicit notion of how long it is prudent to allow children to gain language proficiency before we judge them by the standards set out for the system. During that time, we make allowances and evaluate them by relative rather than absolute standards, accepting progress itself as the measure of success. This is the difference between the standards set by norm-referenced and criterion-referenced assessments . . . But school success is ultimately determined by the absolute (criterion-referenced) standards of the institution. Parents want their children to do well; they do not want them to do well *for an immigrant*.

A large-scale study by Hakuta, Butler, and Witt (2000) examined this question for immigrant children in California. They tested children in two school districts that had high proportions of nonnative English-speaking children. The two districts also differed in their socioeconomic standing, one being more middle-class (district A) than the other (district B). In district A, the more affluent of the two, the study included all the children who had been designated LEP (limited English proficiency) when they entered kindergarten and were between first and sixth grades at the time of testing. This sample included 1,872 children. In district B, a sample of 122 children in grades one, three, and five were selected according to the same criteria, namely, that they entered the school district in kindergarten and were designated LEP. Unfortunately, the testing instruments used in the two districts were not the same so direct comparison between the two districts can be made only cautiously. Still, the design had merits that are largely absent in other evaluations of this type, especially more informal

ones. Notably, assuring that all children entered the school district in kindergarten controlled the age at which English was introduced. If there are any effects of age on the ability and success of learning another language, they were not influencing the results of this study. Minimally, then, the study provides a means of assessing the length of time it took children to reach levels of English that were comparable with native speakers. As a secondary issue, some insight into the role of social class could be gleaned.

The results for the two districts were comparable although the learning latencies were different. In district A, children's test scores were compared with the norm achieved by the native English-speaking students in the same district. By the end of grade four (representing five years of schooling), over 90 percent of the LEP students had achieved oral English proficiency that was comparable to their native-speaking peers. The range of time needed to arrive at this level was two to five years. Academic English proficiency took longer; the 90 percent criterion was not reached until the end of grade six and the range of time needed was four to seven years. The tests used in district B compared children's performance to native speaker norms rather than to classmates in the school district. The LEP children in this board had a more difficult time reaching criterion and the gap between their performance and native speakers actually widened throughout the period examined. In grade three, they were about one year behind the norm in various reading measures, but by grade five they were two years behind. Many differences between the districts, however, prevent a direct comparison of these achievement levels. Some of the factors are economic levels, types of programs available, and nature of the tests used. The general conclusions from this study are that children can reach both age- and peer-referenced norms for English but that it takes time. The first two or three years are a time of rapid growth, but the curve rises at a slower rate for many years after.

Hakuta (1999) states: 'When strict comparisons are made that control for the background factors, children learn English at the same rate regardless of the kinds of programs they are in, i.e., instruction through the native language does not slow down student acquisition of English. It takes most students 2 to 5 years to attain a level of English that does not put them at a disadvantage in regular instruction. Their rate of acquisition of English depends on the level of development of native language – children with strong native language skills learn English rapidly. Motivation to learn English is uniformly high both among parents and the students.'

These empirical findings, despite their limitations, provide considerable support for Cummins's (1979) distinction between *basic interpersonal communicative skills* (BICS) and *cognitive academic language proficiency* (CALP), which has been discussed, developed and criticized extensively in the literature ever since. BICS referred to the kind of language that was used for interpersonal communication in everyday types of situations, whereas CALP referred to the type of language which was required to manage more academic endeavours, and which was required to succeed at school and in the workplace. More recently, Cummins (2000) has used the terms *conversational language proficiency* and *academic language proficiency*. The terms are useful in reflecting the fact that as they go through the formal educational system, children need to acquire very different language skills from those that they are using routinely at home. Cummins's main point is that while children who attend school take approximately two years to achieve conversational language skills,

they take much longer – between five and seven years – to acquire the level of language required for the increasingly demanding academic tasks they will need to perform as they move through the school system.

Cummins was particularly interested in children in circumstantial bilingual situations. This is a substantial group in many countries where there is widespread immigration, such as the US, Canada and Australia, and in many countries in Europe. But it is also a matter of serious concern in many other countries where multilingualism is normal and where the language of education is often different from the language(s) that children learn in their home environment. For example, children growing up in Java, Indonesia, learn Javanese at home, but their education will be in Bahasa Indonesian – the lingua franca of education in Indonesia. Children growing up in New Caledonia learn their indigenous languages at home, but are educated in French. Indigenous children in many countries learn a different language at home from that of their education (e.g. many Indigenous children in Australia do not learn standard English at home – they may learn either a creole, a traditional language or a mixed language – but most will be educated in standard Australian English when they go to school).

Cummins argues that when children begin to attend a school system where the language of education is other than their first language, they will often develop conversational language proficiency relatively quickly. However, academic language proficiency takes much longer to acquire. In other words, children entering school with a first language other than the language of education may rapidly learn the language and be able to communicate quite effectively at a surface level (i.e. they might attain conversational proficiency) but they will not necessarily have the skills to manipulate the more academic language required for long-term success in school. Although Cummins's model does not go uncriticized, it can be useful for evaluating the types of tasks and activities that children in school will encounter and need to be able to manage. The model provides helpful guidelines about both evaluating the language skills required and observing the children's abilities to manipulate these language skills.

In Text B2.3 below, Cummins outlines a framework for assessing language skills, in which he emphasizes the importance of the social context of language use. Cummins points out the importance of recognizing the context for which the framework was developed, arguing that the framework was designed for one particular socio-cultural context – that of schooling. Cummins emphasizes the importance of using it only for evaluation within this specific context.

**J. Cummins (2000) *Language, power and pedagogy: Bilingual children in the crossfire*. Clevedon: Multilingual Matters, pp. 66–71.**

Text B2.3  
J. Cummins

### Cognitive and contextual demands

The framework outlined in Figure B2.3.1 is designed to identify the extent to which students are able to cope successfully with the cognitive and linguistic demands made on them by the social and educational environment in which they are obliged to function in school. These demands are conceptualized within a framework made up of the intersection of two continua, one relating to the range of contextual support available for expressing or receiving meaning and the other relating to the amount of information that must be processed simultaneously or in close succession by the student in order to carry out the activity. While cognitive demands and contextual support are distinguished in the framework, it is not being suggested that these dimensions are independent of each other. In fact, as Frederickson and Cline (1990) point out, increasing contextual support will tend to lessen the cognitive demands – in other words, make tasks easier.

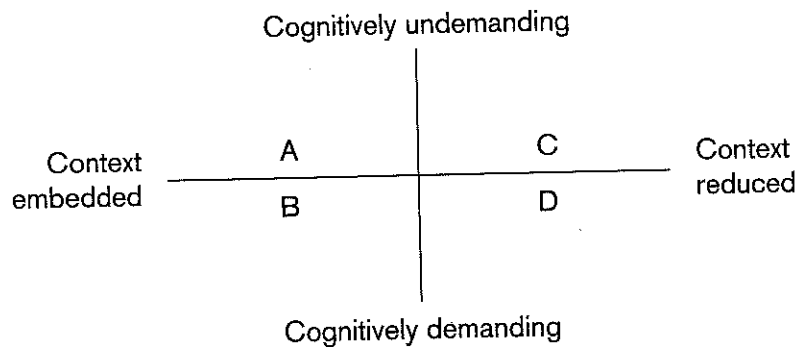


**Figure B2.3.1** Range of contextual support and degree of cognitive involvement in language tasks and activities

### Situating the framework

Before describing the framework, it is important to place it in an appropriate context of interpretation and to define some of the terms associated with it. In the first place, the framework, and the associated conversational/academic language proficiency distinction, focuses only on the sociocultural context of schooling. Thus, we are talking about the nature of language proficiency that is required to function effectively in this particular context. The framework was not intended to have relevance outside of this context.

Related to this is the fact that language proficiency and cognitive functioning can be conceptualized only in relation to particular contexts of use. Just as sociocultural or Vygotskian approaches to literacy emphasize that 'literacy is always socially and culturally situated' (Pérez, 1998a: 4) and cannot be regarded as content-free or context-free, language proficiency and cognitive functioning are similarly embedded in particular contexts of use or *discourses* which are defined by Pérez as the ways in which 'communicative systems are organized within social practices' (1998b: 23). Thus, the social practice of schooling entails certain 'rules of the game' with respect to how communication and language use is typically organized within that context. In short, in the present context the construct of *academic language proficiency* refers not to any

absolute notion of expertise in using language but to the degree to which an individual has access to and expertise in understanding and using the specific kind of language that is employed in educational contexts and is required to complete academic tasks. Drawing on the categories distinguished by Chapelle (1998), academic language proficiency can be defined as the language knowledge together with the associated knowledge of the world and metacognitive strategies necessary to function effectively in the discourse domain of the school.

... [T]his perspective is consistent with an interactionist perspective on language ability 'as the capacity for language use' (Bachman & Cohen, 1998: 18). Current theoretical approaches to the construct of language proficiency have shifted from viewing proficiency as a trait that individuals possess in varying degrees to seeing it as inseparable from the contexts in which it will be manifested. Thus, Bachman and Palmer (1996) use the term 'ability-task' to refer to the unity of ability and task (context) characteristics. Chapelle (1998: 44) similarly notes that a construct such as 'vocabulary size' 'cannot be defined in an absolute sense but instead is a meaningful construct only with reference to a particular context'. Thus, in the context of schooling, discussions of greater or lesser degrees of language proficiency or 'adequacy' of an individual's proficiency refer only to the extent to which the individual's language proficiency (CALP) is functional within the context of typical academic tasks and activities. As noted above and elaborated below, the characteristics of instruction (context) will determine the functionality or 'adequacy' of an individual's proficiency in the language of instruction as much as the degree of proficiency in any absolute sense.

In this regard, it is helpful to introduce the notion of *register* which will be elaborated further in subsequent sections. Register is defined in the *Concise Oxford Dictionary of Linguistics* (Matthews, 1997: 314) as 'a set of features of speech or writing characteristic of a particular type of linguistic activity or a particular group when engaging in it ... journalese is a register different from that in which sermons are delivered, or in which smutty stories are told'. Registers are the linguistic realizations of particular discourse contexts and conventions. Academic language proficiency thus refers to the extent to which an individual has access to and command of the oral and written academic registers of schooling.

In summary, as students progress through the grades, the academic tasks they are required to complete and the linguistic contexts in which they must function become more complex with respect to the registers employed in these contexts. Not only is there an ever-increasing vocabulary and concept load involving words that are rarely encountered in everyday out-of-school contexts but syntactic features (e.g. passive rather than active voice constructions) and discourse conventions (e.g. using cohesive devices effectively in writing) also become increasingly distant from conversational uses of language in non-academic contexts. The framework outlined in Figure B2.3.1 presents an analytic scheme for mapping in a general way how the construct of language proficiency can be conceptualized in terms of the intersections of cognitive (or information-processing) demand and contextual support in academic situations.



### Task B2.3.1



- Think about the kinds of activities children are required to do at school and how these change as children go through the school system. On a piece of paper, label three columns each with one of the following age groups: five, nine and fourteen. In each, list one or two speaking, reading and writing tasks that you might expect a child to encounter at school at each of these ages.
- What are the different language requirements of these tasks?

#### *Description of the framework*

Text B2.3  
J. Cummins

The extremes of the context-embedded/context-reduced continuum are distinguished by the fact that in context-embedded communication the participants can actively negotiate meaning (e.g. by providing feedback that the message has not been understood) and the language is supported by a wide range of meaningful interpersonal and situational cues. Context-reduced communication, on the other hand, relies primarily (or, at the extreme of the continuum, exclusively) on linguistic cues to meaning, and thus successful interpretation of the message depends heavily on knowledge of the language itself. In general, as outlined in the previous section, context-embedded communication is more typical of the everyday world outside the classroom, whereas many of the linguistic demands of the classroom (e.g. manipulating text) reflect communicative activities that are close to the context-reduced end of the continuum [in Figure B2.3.1].

The upper parts of the vertical continuum consist of communicative tasks and activities in which the linguistic tools have become largely automatized and thus require little active cognitive involvement for appropriate performance. At the lower end of the continuum are tasks and activities in which the linguistic tools have not become automatized and thus require active cognitive involvement. Persuading another individual that your point of view is correct, and writing an essay, are examples of Quadrant B and D skills respectively. Casual conversation is a typical Quadrant A activity while examples of Quadrant C are copying notes from the blackboard, filling in worksheets, or other forms of drill and practice activities.

The framework elaborates on the conversational/academic distinction by highlighting important underlying dimensions of conversational and academic communication. Thus, conversational abilities (Quadrant A) often develop relatively quickly among immigrant second language learners because these forms of communication are supported by interpersonal and contextual cues and make relatively few cognitive demands on the individual. Mastery of the academic functions of language (academic registers or Quadrant D), on the other hand, is a more formidable task because such uses require high levels of cognitive involvement and are only minimally supported by contextual or interpersonal cues. Under conditions of high cognitive demand, it is necessary for students to stretch their linguistic resources to the limit to function successfully. In short, the essential aspect of academic language proficiency is the ability to make complex meanings explicit in either oral or written modalities by means of language *itself* rather than by means of contextual or paralinguistic cues (e.g. gestures, intonation etc.).

As students progress through the grades, they are increasingly required to manipulate language in cognitively demanding and context-reduced situations that

Text B2.3  
J. Cummins

differ significantly from everyday conversational interactions. In writing, for example, as Bereiter and Scardamalia (1981) point out, students must learn to continue to produce language without the prompting that comes from a conversational partner and they must plan large units of discourse, and organize them coherently, rather than planning only what will be said next.



### Task B2.3.2

- Take the list of tasks you developed in Task B2.3.1. Try to classify each of these tasks into one of the quadrants in the figure. Put the age group the task is for in brackets after the task (e.g. copying sentences (5)); composition about recent holiday (9); written science report (14). How well do you find the tasks fit neatly into the quadrants? What, if any, difficulties do you encounter in trying to allocate the tasks to the different quadrants? To what extent do the tasks change with age from context embedded to context reduced, and cognitively undemanding to cognitively demanding?

Text B2.3  
J. Cummins

Evidence for the relevance and usefulness of the conversational/academic language distinction comes from observations made by Carolyn Vincent in an ethnographic study of a program serving second generation Salvadorean students in Washington DC:

All of the children in this study began school in an English-speaking environment and within their first two or three years attained conversational ability in English that teachers would regard as native-like. This is largely deceptive. The children seem to have much greater English proficiency than they actually do because their spoken English has no accent and they are able to converse on a few everyday, frequently discussed subjects. Academic language is frequently lacking. Teachers actually spend very little time talking with individual children and tend to interpret a small sample of speech as evidence of full English proficiency. However, as the children themselves look back on their language development they see how the language used in the classroom was difficult for them, and how long it took them to acquire English. (1996:195)

In this respect, Vincent notes on the basis of her classroom observations: 'It is clear that student achievement is promoted by instructional practices such as cooperative learning, the use of manipulatives, and project-based lessons' (1996: 201). These are activities that place a significant emphasis on Quadrant B insofar as they tend to be cognitively demanding but contextually supported.

Pauline Gibbons has similarly expressed the difference between the everyday language of face-to-face interaction and the language of schooling in outlining the distinction between what she terms *playground language* and *classroom language*:

This playground language includes the language which enables children to make friends, join in games and take part in a variety of day-to-day activities that develop and maintain social contacts. It usually occurs in face-to-face contact, and is thus highly dependent on the physical and visual context, and on gesture

Text B2.3  
J. Cummins

and body language. Fluency with this kind of language is an important part of language development; without it a child is isolated from the normal social life of the playground . . .

But playground language is very different from the language that teachers use in the classroom, and from the language that we expect children to learn to use.

The language of the playground is not the language associated with learning in mathematics, or social studies, or science. The playground situation does not normally offer children the opportunity to use such language as: *if we increase the angle by five degrees, we could cut the circumference into equal parts*. Nor does it normally require the language associated with the higher order thinking skills, such as hypothesizing, evaluating, inferring, generalizing, predicting or classifying. Yet these are the language functions which are related to learning and the development of cognition; they occur in all areas of the curriculum, and without them a child's potential in academic areas cannot be realized. (1991: 3)

Thus, the context-embedded/context-reduced distinction is not one between oral and written language. Within the framework, the dimensions of contextual embeddedness and cognitive demand are distinguished because some context-embedded activities are clearly just as cognitively demanding as context-reduced activities. For example, an intense intellectual discussion with one or two other people is likely to require at least as much cognitive processing as writing an essay on the same topic. Similarly, writing an e-mail message to a close friend is, in many respects, more context-embedded than giving a lecture to a large group of people.

It follows that cognitive academic language proficiency is not synonymous with literacy. It is manifested as much in oral interactions in academic contexts as in written interactions. For example, a classroom or small group discussion of the social consequences of industrial pollution will draw on participants' familiarity with features of academic registers (e.g. how to express cause-effect relationships linguistically) and will reveal the depth and richness of their understanding of words and concepts. In Aitchison's (1994) terms, the extent to which students have built extensive semantic networks will be evident in the way they use and relate words and concepts in such discussions. Thus, words are not just known or unknown; there is a continuum with respect to depth of vocabulary/concept knowledge (Paribakht & Wesche, 1997; Verhallen & Schoonen, 1998). One of the major functions of schooling is to deepen and broaden students' knowledge of words and their meanings and to develop what Norah McWilliam (1998) calls *semantic agility*. Oral classroom discussions do not involve reading and writing directly, but they do reflect the degree of students' access to and command of literate or academic registers of language. This is why CALP can be defined as *expertise in understanding and using literacy-related aspects of language*.

Skourtou has provided a particularly clear description of how the processing of experience through language transforms experience itself and forms the basis of literacy:

It seems to me that the entire process of language development both starts and ends with experience. This implies that we start with a concrete experience, process it through language and arrive at a new experience. In such a manner, we develop the main features of literacy, namely the ability to reconstruct the world through symbols, thus creating new experiences. Creating experiences through language, using the logic of literacy, whether speaking or writing, means



Text B2.3

J. Cummins

that once we are confronted with a real context, we are able to add our own contexts, our own images of the world. (1995: 27)

Highlighted here is the fact that *context* is both internal and external, representing two sides of the same experiential coin; furthermore, the development of literate modes of thought cannot be separated from these experiential contexts.

It should now be clear that this model Cummins presented has major implications for pedagogy as well as for the assessment of language ability in the school context. Its value, as Cummins points out, lies in showing how learners can move from context-embedded and cognitively undemanding tasks to more cognitively demanding tasks, which remain none the less context-embedded. From this point learners can move into more context-reduced tasks, which may vary between those which are more cognitively demanding and those which are less cognitively demanding. He provides an example, taken from the work of Gibbons (1995), in which a science class began with the students working in small groups. In this context they were able to use context-embedded language. They then moved to a context where the teacher was standing in front of the class and required the children to describe and explain the activities they had been carrying out in their small groups. This allowed the children to be led into the use of more decontextualized language, but with material with which they were very familiar from working in groups. After this, the children were asked to write up their reports, using some of the language which had earlier been introduced by the teacher in a discussion which led up to the written work. Cummins uses the example of the teacher introducing the word 'repel' to replace the children's more familiar 'push away' (2000: 71).

The model presented by Cummins provides us with a useful tool both to understand the range of language children need for effective schooling and also to evaluate the extent to which children are able to manipulate different aspects of this language, as well as to manage the pedagogical content of the classroom.

### Summary

As we have seen in this unit, the measurement of language is never easy. As it is a latent (or invisible) trait, we need firstly to define exactly what language is so that we are able to know what to measure. In addition to this, the complexities of defining bilingualism itself (see Unit A1) mean that this will always be a difficult task. There is an increasing awareness in the measurement literature of the problems of bias with respect to the testing of bilingual children who come from linguistically and culturally diverse populations (see Unit A4). Yet for children to be able to participate fully in both education and later in the workforce, it is crucial to be able to identify potential developmental delays in language accurately, and to be able to assess children in the educational system in as fair and unbiased a way as we can.