

## Test translation and adaptation in public education in the USA

Charles W. Stansfield *Second Language Testing, Inc., N. Bethesda, MD*

The use of translated versions of formal educational assessments is becoming more common in the United States of America (USA). This is due to two converging factors:

- legislation passed during the Clinton and Bush administrations;
  - an influx of non-English-speaking immigrants, especially from Latin America.
- Against this background, this article discusses theoretical and practical issues pertaining to the translation or adaptation of educational assessments in the USA. These issues include the role of language proficiency and academic background in performance on standards-based achievement tests in different languages, factors affecting the decision whether to translate tests, translation methods and procedures, the degree to which translated tests are used in the USA, current laws influencing the use of translation, and the role that language testing specialists may play in this arena.

### I Assessment and education reform

For the past 25 years, there has been a movement in the USA to establish greater accountability for student achievement among educators. The movement has been aided by the passage of the Improving America's Schools Act (IASA) (USA Congress, 1994b). IASA and the Goals 2000: Educate America Act (USA Congress, 1994a) are based on the premise that all children enrolled in public schools must be held to high standards and assessed accordingly with appropriate instrumentation. In the United States of America (USA), public education is the responsibility of the states, rather than the federal government. However, the federal government may provide funding to states for various kinds of educational programs. The federal government provides financial support to states under IASA, which requires states to address the issues of standards and accountability.

An important requirement of IASA was that states develop and adopt a set of challenging content standards for the acquisition of

---

Address for correspondence: Charles Stansfield, Second Language Testing, Inc., 10713 Mist Haven Terrace, N. Bethesda, MD 20852, USA; email: cstansfield@2lti.com

*Language Testing* 2003 20 (2) 188–206 10.1191/0265532203lt252oa © 2003 Arnold

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82

reading and mathematics by the 1997–98 school year. These standards must specify what all children must know and be able to do in these disciplines in at least three grade levels (for example, grades 4, 8, and 10).

Another important requirement of IASA was that states develop and adopt assessments based on these content standards by the 2000–01 academic year. As of January 2001, 40 states had developed their standards-based assessment system. In most of these states, the assessment system goes beyond the testing of reading and mathematics. Science, social studies, health, and other subjects may also be tested. The IASA requires that all children be included in these assessments, including limited-English-proficient (LEP) children. LEP children are normally included in these programs through the use of accommodations and modifications of the assessment or assessment conditions. These accommodations allow and encourage the use of assessments in the student’s native language. As a result, a number of states now routinely translate their assessments to one or more non-English languages.

This article discusses test translation in the context of standards-based assessments used in elementary and secondary education in the USA. The article discusses practical issues pertaining to the translation or adaptation of educational assessments, methods and procedures that are used, the degree to which translated tests are used, current laws influencing the use of translation, and the role that language testing specialists may play in this arena.

## 83 **II The language proficiency issue**

84 Over 20 years ago, Oller and others (Oller and Perkins, 1978; Gunnarsson, 1978) argued convincingly that tests of intelligence, personality, attitudes, and achievement ask very similar questions and test primarily English language skills for an examinee who is not fully proficient in English. Studies of college admissions tests have shown that English language is a moderating variable on aptitude test scores (Alderman, 1982). That is, an examinee’s score is largely dependent on his or her proficiency in the language of the test, at least up to the point where the examinee’s language proficiency begins to approximate the language proficiency of the educated native speaker. This shows that, regardless of the construct claimed to be measured by the test, for the nonnative speaker the test is primarily a test of language proficiency.

87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99

The issue of language proficiency is addressed in the new Standards for Educational and Psychological Testing (Joint Committee, 1999: 98). Standard 9.3 states, ;When testing an examinee proficient in two

or more languages for which the test is available, the examinee's relative language proficiencies should be determined. The test generally should be administered in the test-taker's more proficient language, unless proficiency in the less proficient language is part of the assessment.' Concern for the need to determine if an examinee has adequate proficiency in English to participate meaningfully in the regular testing program (the English version of the tests) has led some states to develop their own ESL assessments specifically for this purpose. To date, only Illinois has completed such a test, but the states of Texas, California, Massachusetts, and Minnesota are currently working on them. Illinois developed the Illinois Measure of Annual Growth in English (IMAGE) (Woods Chapman, 1996). One purpose of the IMAGE is to determine if a language minority student should participate in the regular testing program, if the student has attended state schools for fewer than three years. Minnesota has completed development of the Test of Emerging Academic English for the same purpose, but for a variety of reasons the state has not yet implemented the test.

The importance of language proficiency in test performance must be considered when evaluating the suitability of a translated or adapted test for a given examinee. Since 1975, nonnative English speaking students in the USA have been classified into three groups: non-English proficient (NEP), limited English proficient (LEP), and fully English proficient (FEP). The NEP and LEP designations are sometimes called LOTE (language other than English) in other English-speaking countries. A translated/adapted test is normally more appropriate for the NEP examinee, because the regular English version would not be accessible to the student. At first glance, a translated/adapted test may appear appropriate for the LEP student. However, this is not always the case. The appropriateness of either the regular or the translated/adapted versions depends on a number of factors.

The most important factor is the language proficiency of the student in the two languages for which the test is available. Language proficiency varies according to the domain in which it might be used, and the kind of language skill required by the situation. Cummins (1984) has distinguished between language use that is cognitively undemanding and that which is cognitively demanding. He also contrasts language situations that are rich in supporting context with those that are decontextualized. These distinctions apply to the kind of language skills called for when taking achievement tests. That is, the tasks associated with a test are often cognitively demanding and both the tasks and the language occur in situations of reduced contextual support.

That is, a given individual may have very complete proficiency in one domain of language use and only partial proficiency in another. For example, the examinee may have good speaking proficiency, but poor reading proficiency in the native language. Thus, if the LEP examinee does not have the kind of native language proficiency called for by the test, then the native language version will not produce a score that reflects the examinee's true ability. Indeed, the regular version of the test may even provide a more accurate score, if the examinee has stronger reading proficiency in the language of the regular version.

### III Academic background

Another factor is the examinee's academic background in the language of the test. While an LEP examinee may have greater proficiency in the language of the translated version of the test, most of his or her knowledge of the subject matter may have been acquired in an English-medium instructional setting. This instruction will develop the examinee's knowledge of content-relevant vocabulary and expression in English, but not in the native language. Thus, through English-medium instruction in the subject matter, the examinee may acquire the ability to interact with an achievement test in English, but this ability may not transfer readily to the examinee's native language. Thus, decoding the test stimuli and items in the stronger language may be more difficult for the examinee, and the regular version may provide a more accurate measure of true ability in the construct being assessed. Just as studies have shown that background knowledge affects performance on language tests (Alderson and Urquhart, 1983; 1985; Hale, 1988), so can the acquisition of academic knowledge in a language alter the appropriateness of tests of that discipline in two different languages for a bilingual examinee. Those who are considering using translated or adapted tests for LEP examinees need to be sensitive in judging the role of language proficiency in the discipline and the role of content knowledge in each test version when determining which version will be more appropriate.

### IV Translation as an accommodation

Accommodation is a broad term that refers to any specific change to the testing situation (e.g., presentation format, response format, setting, and the timing/scheduling of tests). Test program administrators have created Braille or large print versions of tests for the visually impaired for many years. These versions are accommodations in the

presentation format. The utilization of accommodations for LEP students is a means of enabling them to demonstrate their academic knowledge despite their lack of English proficiency. The purpose of an accommodation is to ‘level the playing field’ for examinees with special needs. In this sense, a translated or adapted test is a test accommodation. Because translation and adaptation are considered accommodations, it is natural for test program administrators to consider them, just as they consider other accommodations for students or examinees with special needs.

## **V Issues to be considered**

### *1 Cost efficiency*

The cost of translating or adapting a test is a major issue in deciding whether to provide a native language version of a test. Test translation is far more expensive than the translation of other kinds of documents, because of the many additional steps and extensive reviews that must be built into the process. If the test is to be adapted, then the cost can approach the cost of developing a completely new test.

### *2 Numbers*

A central issue in considering the cost efficiency of a translated or adapted test is the number of students that may benefit from it. Brute cost and numbers have restricted the creation of translated versions in languages other than Spanish in most states that translate tests in the USA. Some states develop policies that allow for native language versions of tests, and then do not create them after determining the cost and number of students who would be served. Other states have developed non-English versions of their assessments, and then learned through experience that fewer students than expected opt to take them instead of the English version. Thus, they cease to develop non-English versions of such tests. For example, in 1995, the state of Rhode Island developed translated versions of its state assessments in all subject areas and at multiple grade levels in Spanish, Portuguese, Lao, and Khmer. When only a handful of non-LEP students in languages other than Spanish took the non-English versions, the state discontinued development of non-English versions in languages other than Spanish.

As of yet, there is no commonly used minimum number for determining whether it is cost efficient to develop a non-English version of a test. However, in its district assessment plan, the school district of Philadelphia specifies that it will translate a district-wide

assessment to a language if 50 or more students can be identified for whom the translated version would be the most appropriate measure.

### 3 Attitudes and politics

A third influence on the decision whether to offer a non-English version of a state assessment is language attitude. The attitude may have to do with a belief in the primacy of English in the state and in its educational system. Some state education officials may feel that non-English tests set a bad precedent, while others may fear that offering such tests will generate public criticism of the assessment program. Conversely, in states where a specific ethnic group wields considerable political influence in state government, the language of that group is more likely to appear on state assessments. Spolsky (1997) states that tests have always been used as a means of political and social control. While attitudes towards English and non-English languages play a role in some programmatic decisions, such considerations are less influential than matters of cost and numbers.

## VI The translation process

While there is much that could be said about the test translation/adaptation process, we limit this discussion to a procedure about which there is considerable division and conflicting literature. This procedure is known as 'back translation.' Back translation has been proposed as an improvement over the traditional forward translation, which is a translation from the source language into the target language. A perceived problem with forward translation is that the test program administrator does not know if the target language document is equivalent in meaning, unless he or she knows the target language or languages.

A cross-cultural psychologist, Brislin, has written extensively about back translation (1970; 1976; 1986; Brislin *et al.*, 1973). Brislin recommends back translation as a method that produces a high quality translation of a test instrument. The instrument can then be used to examine traits and constructs across cultures. Following his lead, a number of other authors (Werner and Campbell, 1970; Bernard, 1988; 1994; McKay *et al.*, 1996) have written about it as well. Back translation as described in the literature on cross-cultural research involves asking one or more bilinguals to translate the original test to the target language, and then having a different bilingual translate it back to the source language, in this case English. The two English (i.e., source language) versions are then compared, and points of disagreement are

used to identify problems in the initial forward translation. The forward translation is then corrected. Sometimes the process of creating a back translation and comparing it with the source document is repeated (Marin and Marin, 1991). Back translation is viewed in the test translation literature as a method for drafting, reviewing, and revising a translation. In fact, the purpose of the back translation is to identify and correct errors in the forward translation.

In this author's opinion, there are a number of problems in relying on back translation to examine the quality of the translated document. First, the lack of agreement between the original document and the back translation may be due to problems with the back translation, not to problems with the forward translation. That is, the back translation is as likely to contain translation errors or infelicities as is the forward translation. Then, the test program administrator is left with two translations (the forward and the backward) and no verification of the quality of either. Once the back translation is completed, the focus of work becomes a comparison of the two English documents (the original test and the back translated version), as one searches for points of inconsistency. One must search the two translations to determine the reason for the inconsistency. If the reason relates to the forward translation, only then does one begin to consider possible solutions to the problem.

Secondly, when a translator knows that the initial forward translation will be validated by a back translation procedure, this influences the nature of the forward translation. By producing a very literal forward translation, the translator can ensure that the back translation will produce a document that is highly similar or identical to the source document. However, the literal forward translation may represent stilted rather than natural expression in the target language. As a quality control procedure, back translation essentially ignores such stilted language, even though the resulting forward translation may be difficult to read.

Thirdly, because back translation can send false messages about the forward translation, and because it encourages a literal and unnatural forward translation, it can result in a considerable waste of time and money. The money expended on back translation can be better spent on other aspects of the testing program.

Fourthly, the literature on back translation comes from the field of cross-cultural psychology. Researchers in this field tend to rely on the use of bilinguals rather than professional translators to do the translation. Professional translators are themselves bilingual, and almost always translate into their native language. In addition, they normally have outstanding writing skills in the target language. Selecting someone to do a translation merely because they claim to

1  
2  
305 be bilingual is not likely to produce a high quality equivalent docu-  
306 ment in the target language. This does not mean that all professional  
307 translators are equal in competency or that they have the proper train-  
308 ing to translate a test. Professional translators differ as much in com-  
309 petency as professionals in any other field. However, as a starting  
310 point, test contractors should make use of translation professionals  
311 with appropriate specialized knowledge and skills in the subject  
312 matter of the test. With those basic credentials, they will need to be  
313 taught basic principles of item writing.

314 It is not our intention here to say that back translation is a flawed  
315 procedure. Indeed, it is a method of quality control that can be useful  
316 to the test program manager who does not have confidence in the  
317 translation reviewers who are available. Normally, however, other  
318 methods – e.g., successive reviews and revisions of a forward trans-  
319 lation by other professional translators and by representatives of the  
320 examinee population – are available to identify and correct problems  
321 with the forward translation. For this reason, professional translation  
322 agencies do not use back translation as a quality control procedure,  
323 unless specifically requested to do so by their clients. Instead, they  
324 rely on successive review and revision of the forward translation by  
325 reviewers of known competence as translators.

326 In this sense, the translation of test items is an iterative process  
327 much like item writing. In item writing, an item is drafted, then  
328 reviewed and revised. The process is repeated several times until  
329 additional reviews produce suggestions that have already been con-  
330 sidered or until no new useful suggestions are made. The same hap-  
331 pens in the review and revision of test stimuli and test items.

332 In state assessment, some states have used back translation. Minne-  
333 sota experimented with back translation to obtain feedback on its  
334 translation on one of several state assessments to Spanish (Liu *et al.*,  
335 1999). A full year was devoted to the translation and pilot testing of  
336 the test. In the real world of state assessment, only two to four weeks  
337 are typically available to translate, review, revise, typeset, proof, and  
338 correct a test. While the back translation proved successful in Minne-  
339 sota in that it helped produce an equivalent translation, Liu *et al.*  
340 noted the high cost of the translation. Successive iterations of forward  
341 translation and revision would work just as well, and the translation  
342 equivalence of the two versions can be assured more quickly and at  
343 less cost.

## 344 VII Translation vs. adaptation

345 A translated test is one where the same exact content is rendered into  
346 a non-English language (i.e., non source language). The standard test

and translated test then differ only in language, not in content. An adaptation is a modified version of the standard assessment. Due to the nature of some tests and their items, adaptation is required in order for the standard test to be appropriately rendered into a non-English language. Adaptation may involve removing some items and replacing them with others that are more appropriate for the native language or more valid for the examinee population or for the language of the new test.

An example of why adaptation is necessary may be useful. If a professional or trade association were to create a certification of competency for auto mechanics in the USA, the test would contain several core sections, such as a section dealing with the theory of the internal combustion engine. The test would also contain a section testing applied matters, such as knowledge of the parts and repair procedures associated with cars manufactured in the USA, in Japan, Korea, and Europe. If the association decided to expand its certification program to France, and to certify mechanics who work in France, certain changes in the test would be necessary. While the core section on internal combustion would still be necessary, the French mechanic would probably have a different kind of applied knowledge, reflecting the kinds of automobiles he or she is likely to encounter. Therefore, in addition to translating the test to French, some items would have to be dropped and others would have to be added so that the test reflected the competencies needed by a mechanic in France. The new test would be considered an 'adaptation' of the original test.

While the distinction between translated and adapted tests is important, even translated tests normally require minor adjustments to accommodate the language of the non-English version. For example, a mathematics or science test translated to Spanish will reverse the use of commas and periods in the translated version (the number 10,215.64 is written 10.215,64 in Spanish).

Adaptation affects our ability to compare scores on the standard and adapted versions. The change in test content raises validity concerns, especially if a substantial number of items are changed. As a result, it becomes necessary to demonstrate the equivalence of the constructs measured by the standard and adapted instruments. Because this process is long and expensive, adaptation is rarely used in state assessments in the US. Instead, tests whose validity and comparability may change if translated or adapted are normally not translated or adapted at all. The most common example of this in state assessment is the decision not to translate tests of Language Arts. Such tests are measures of reading comprehension skills, spelling, and grammar in English. If a test of English Language Arts were rendered into Spanish, it would have to be adapted, since a large number of

items would have to be modified or replaced in order to create a valid test of Language Arts in Spanish. While undoubtedly some items could be directly translated to Spanish (certain reading comprehension and grammar items, i.e., those addressing subject–verb agreement), the resulting test would constitute a major adaptation of the original.

As indicated above, and in other articles in this issue, adaptation raises concerns about the comparability of the constructs assessed by the original and translated tests. Translation raises concerns about the need for statistical adjustments to eliminate any bias in the test score that may result from the translation. However, it is difficult to come up with an appropriate linking design, since English-speaking and LEP/NEP examinees represent two different populations with different educational backgrounds. As a result, differences in item difficulty and in mean test scores may be due to differences in the two test populations in the mastery of the educational content and constructs assessed. Elder (1997) has noted that group differences may be explained in two ways:

- There is a real difference in the ability being tested.
- There are confounding variables within the test that systematically mask or distort the ability being tested.

Hambleton (1993) and Sireci (1997) have also noted the limitations in the linking designs that have been proposed. Because of the difficulty of collecting data that would permit test program administrators to know whether mean differences are due to differences in language of the test, rather than differences in content mastery in the two test populations, translated versions of state assessments are normally scored on the same scale as the English version.

### VIII Beyond adaptation: a different test

As one begins to plan for assessments in the native language, the issue of validity arises immediately. It can be argued that because the examinee population is different in language, culture, and educational background, a whole new test should be developed. However, in the context of standards-based state assessment, the new test would face the problem of demonstrating equivalent reliability and validity in assessing the state standards. Tests that are mostly or completely different in content are considered ‘alternate assessments.’ While the use of alternate assessments for special populations is growing, such assessments will always generate concerns about their validity. That is, do the alternate assessments really assess the content in the same way and do the scores mean the same thing or reflect the same standards as scores obtained on the regular assessments?

**IX The test booklet**

There are two basic options when administering a translation or adaptation of a test in another language. One is to produce test booklets in both languages and then allow either the teacher or the student to determine which booklet should be used by the examinee. Another is to produce the test booklet in a format that puts the two languages on facing pages. A variant of this option uses parallel columns; for example, with the left column in English and the right column in Spanish. Other variations are also possible, including having students take the test in both languages and giving the student the higher of the two scores. Related to the issue is whether additional time should be allotted to examinees who take the test in the parallel column or facing pages format.

In 1997, Stansfield (1998) conducted an informal survey for the Massachusetts Department of Education concerning test booklet formats using the internet to collect descriptions of practice around the world. The results showed that different approaches are used in different countries and that test program administrators are confident their approach is working well, regardless of which approach is used. However, these same administrators believe that competing test booklet formats are flawed in ways that may affect the validity of the test score. For example, test program administrators who do not use a bilingual test booklet feel strongly that this format is not authentic because students are not typically presented with bilingual texts to process. They also feel that the format changes the cognitive processes tapped in taking the test, and therefore the validity of the score.

Perhaps most interesting was a report from Mexico regarding translated tests in Maya and Spanish. Students are asked to take the test in both languages. The test program administrators feel that if students are asked to choose the language of their test, they will be influenced by the expectation that they take the test in the dominant language of the larger society. In order to avoid any order effect, they are given the higher of the two scores.

In order to evaluate the effects of different test booklet formats, the Massachusetts Department of Education used two different formats when field testing its state assessments in the fall of 1997. Spanish-speaking LEP students were either given a bilingual test booklet, with Spanish and English on facing pages, or they were given a Spanish-only booklet, but were also told they could see the English booklet if they wished to during the test. Each student's teacher made the decision as to which test booklet format the student would receive, and the teacher administered the tests.

After the field test administration, three staff of the Massachusetts Department of Education interviewed 97 students and 17 teachers. The result indicated that students who received bilingual test booklets relied mostly on the Spanish version of the items. However, in some cases they also read the English version and felt that they gleaned some additional meaning from this version. As a result, it was decided that in the future, only a bilingual version of the test would be printed in addition to the English version.

A similar study in Minnesota (Liu *et al.*, 1999), involved a bilingual version of the state reading assessment. In interviews conducted after field-testing, project staff found that most students relied on only a single language version, but some students did refer to both versions. These students reacted positively to being able to see the test on facing pages, while none of the others reported finding the format distracting or confusing.

A recent study (Garcia *et al.*, 2000) conducted in preparation for a national assessment of mathematics investigated how groups of students with different background characteristics react to bilingual test booklets. After taking an eighth grade mathematics test via a bilingual test booklet, students were interviewed by the researchers. In the interviews it was determined that native Spanish speakers with more than three years instruction in the USA focused on the English pages the majority of the time. Those with fewer than three years focused on the Spanish pages the majority of the time. Twelve of the 70 students with more than three years instruction in the USA used the Spanish pages to check their understanding of the English item or to look for the meaning of specific words and phrases (e.g., bisector). Of 181 bilingual students who took the assessment with a bilingual test booklet, 85% rated the bilingual format as useful or very useful for them, in comparison with an English test booklet. Regression analyses with other variables showed that the dual language format did not differ in the validity of the assessment, when compared to similar analysis of response data obtained with an English test booklet.

The bilingual format is also common in multinational computer-based testing programs in the occupational testing arena. Thomson Prometric, a major provider of computer-based certification tests, often allows examinees to see the item in both languages, although this decision is left up to the certification agency. When both languages are allowed, the examinee can tile (display) the two screens and view them simultaneously, or the examinee can cascade them (impose one over another), and even size them on the screen. Although Thomson has not conducted research on whether examinees view this as

2  
 519 advantageous, the company believes that examinees like being able  
 520 to access items in two languages.

521 Experience has shown that not all students who are eligible to  
 522 receive a translated test will choose it over the standard test in  
 523 English. Often, such students are not fully literate in their native language.  
 524 Because of this, bilingual test booklets are often used  
 525 (Stansfield, 1997; Stansfield and Kahl, 1998; Liu *et al.*, 1999).

526 Currently, in the USA, most states allow the teacher to determine  
 527 the language of the test booklet that the student will receive, and in  
 528 some cases the student is given two test booklets, one in each language.  
 529 However, the bilingual format is gaining acceptance in state  
 530 assessment programs. When considered as a potential test accommodation,  
 531 there is a growing belief that it ‘does no harm,’ while at the  
 532 same time relieving the examinee and the test administrator from having  
 533 to decide the language of the test. Since many examinees who  
 534 take translated tests have some degree of bilingualism, making available  
 535 the test in both languages may reduce construct-irrelevant variance  
 536 due to the influence of test language for such examinees.

537 **X Typesetting: the forgotten obstacle**

538 Often, when deciding whether to translate or adapt a test to the  
 539 student’s native language, educators fail to consider the cost and difficulty  
 540 of typesetting the test in that language. Very few professional desktop  
 541 publishing specialists in the language may be available, and the software  
 542 they use may not be the same software used for typesetting the English  
 543 version of the test. Typesetting and reviewing of the typeset version of  
 544 the test can be as expensive as the translation itself, particularly in the  
 545 case of Asian and other languages with non-Roman writing systems. Adding  
 546 to the cost of typesetting are graphics, which may have to be completely  
 547 recreated in the new software for non-English versions, particularly if  
 548 the graphic images imported from the English version contain labels or text  
 549 in English within the graphic. Often, such images cannot be modified  
 550 because the translated text will not fit within the allocated space in the  
 551 graphic.

552 **XI Current policies on test translation in the USA**

553 A recent study of 1998–99 state assessment policies (Rivera *et al.*,  
 554 2000), identified three states (Delaware, Kentucky, and Maine) whose  
 555 policies allow translation of all components of the state assessment  
 556 system to be translated. These are usually direct translations, rather  
 557 than adaptations that involve significant alterations of the original test.  
 558 Eleven states (Colorado, Massachusetts, Minnesota, New Mexico,

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88

New York, Oregon, Rhode Island, Texas, Utah, Vermont, Wyoming) had policies that allow translation of some components (tests) within the system. New Mexico, New York, and Texas have allowed the translation of tests for some time. In most cases, the above 11 states allow translation of all components except Reading, Writing, Language Arts, or other measures of English skills, since tests of English skills would not be directly translatable. The fact that these states have a policy that allows for translation does not mean that a translated version is in fact developed. Cost, numbers, technical concerns, and political concerns may impede the development of translated versions in a given year. These same issues probably explain why in 1998–99, test translation was prohibited in six states. These states are Illinois, Indiana, New Jersey, Pennsylvania, Washington, DC, and West Virginia.

Among the states that produce and offer a formal translation of the assessments, currently only Minnesota and New York translate their assessments into languages other than Spanish. Minnesota also translates its assessments into Vietnamese and Hmong; New York also translates them into Chinese, Haitian Creole, Korean, and Russian.

The author is directing a study of state test translation policies and practices in place during the 2000–01 academic year (Stansfield *et al.*, in progress). Preliminary results indicate that in the two years subsequent to the 1998–99 school year, two more states established formal policies allowing for the translation of their state assessments. In 2000–01, 13 states had such policies. In addition, six states now specify the qualifications of translators in their assessment policy. The number of states prohibiting the translation of their assessments decreased from six to four. These preliminary results confirm a slow but steady growth in the provision of translated assessments in the USA.

## 589 **XII Federal legislation**

590  
591  
592  
593  
594  
595  
596  
597  
598  
599

Although currently less than a third of the states in the USA allow tests to be given in students' native language, that situation is changing rapidly. That is because the final requirements of the Improving America's Schools Act (IASA) of 1994 went into effect during the 2000–01 school year. By Spring 2001, all states receiving federal education funds are required to have an assessment system in place that meets a set of requirements designed to ensure the inclusion of English language learners in the state assessment system. Regarding translation, Section 1111(b)(3)(F) of Title I of the IASA says that state assessments shall provide for:

1

2  
3

1                   202   *Test translation and adaptation in public education*

2  
600                   the inclusion of limited English proficient students who shall be assessed, to  
601                   the extent practicable, in the language and form most likely to yield accurate  
602                   and reliable information on what such students know and can do, to determine  
603                   such student's mastery of skills in subjects other than English.  
604

607                   Section 1111(b)(5) requires that 'each state identify the languages  
608                   that are present in the participating student population and indicate  
609                   the languages for which yearly student assessments are not available  
610                   and needed.' Moreover, an Interpretive Guidance document for states  
611                   issued by the USA Department of Education (Office of Elementary  
612                   and Secondary Education, 1996) to assist in the implementation of  
613                   IASA specifies: 'If native-language assessment is practicable, and if  
614                   it is the form of assessment most likely to yield valid results, then a  
615                   State must utilize such assessments.' Summary Guidance  
616                   (regulations) issued by the USA Department of Education in April  
617                   2000 (Office of Elementary and Secondary Education, 2000) states  
618                   that each state must ensure consistency and appropriateness in the use  
619                   of test accommodations through monitoring and data collection. In  
620                   this area of data collection, states have progressed slowly. As a result,  
621                   there has been little research on the psychometric qualities of  
622                   translated or adapted assessments used in states.

623                   The requirements of IASA, and the subsequent guidance docu-  
624                   ments issued by the USA Department of Education, stimulated  
625                   considerable interest in test translation as a means of including LEP  
626                   students in state assessment systems. As the 2000 presidential elec-  
627                   tions approached, some states delayed moving to translate their  
628                   assessments, on the chance that a change in administration would  
629                   result in a change in policies. President Bush assumed control of the  
630                   government in January 2001, and his administration has continued  
631                   to support standards and assessments, including the requirements for  
632                   assessments in students' native languages (i.e., translated  
633                   assessments). In fact, the IASA requirements were reinforced by the  
634                   2001 reauthorization of the Elementary and Secondary Education Act  
635                   (ESEA), which was renamed the No Child Left Behind Act  
636                   (NCLBA) of 2002. Section 1111(3)(C) requires states to test limited  
637                   English proficient students with:

638                   assessments in the language and form most likely to yield accurate data on  
639                   what such students know and can do in academic content areas, until such  
640                   students have achieved English language proficiency.  
641

644                   The use of tests in a language other than English is permitted for a  
645                   period of three years, but NCLBA allows local school districts to  
646                   extend the testing in a non-English language for an additional two  
647                   years if the student's lack of English proficiency would impede the  
648                   tests in English from yielding valid and reliable results (US Con-  
649                   gress, 2002).

Moreover, a summary guidance document for states recently issued by the USA Department of Education requires each state to develop a plan for complying with the requirements of NCLBA. Among other specifics, each state's plan must 'Indicate the languages for which yearly student academic assessments are not available and are needed' (Office of Elementary and Secondary Education, 2002).

Given the above requirements, native language versions of state assessments will become increasingly common. This will increase the need to ensure that such versions are prepared competently, that appropriate students are selected to take them, and that appropriate interpretations are made of students' scores.

### **XIII What do language testing specialists have to contribute to this area?**

Language testing specialists participate in two broad communities of scholars: applied linguistics and educational measurement. Among applied linguists, their knowledge of measurement often distinguishes them. Among measurement specialists, their knowledge of language matters and their application to the real world distinguishes them. As an area of professional endeavor, test translation has been most seriously investigated by cross-cultural psychologists and by measurement specialists. While each group has much to contribute, neither brings all the background needed to understand the myriad number of issues that can arise in a test translation project. An understanding of the field of translation and the translation profession is very helpful and an understanding of the role of language in test performance is essential. These are areas of understanding that neither the cross-cultural psychologist nor the measurement specialist typically control. The language testing specialist can bring a thorough understanding of the role of language in test performance to the endeavor. As such, he or she can contribute to a more informed discussion of issues affecting examinees. Such issues may include but are not limited to:

- determining if a translated or adapted version of a test is likely to provide valid measurement for an examinee of a certain native language background or with a certain language proficiency profile;
- determining if a translated or adapted version of a test is likely to provide valid measurement for a population with similar language background characteristics;
- determining whether an examinee has adequate English language skills to take the test in English;
- determining if the nonnative English speaking subjects participating in a linking study are appropriate for the purposes of the study.

Language testing specialists would do well to learn more about test translation and adaptation and to participate in such projects. By doing so, they can contribute to the measurement community and to the communities of nonnative speakers that they may have previously worked with in projects that involve only language. Unlike other applied linguists, language testing specialists are comfortable in most discussions among educational measurement professionals. Their grounding in a particular subject area (language) allows them to function comfortably with teachers in a wide variety of disciplines. Not only do they understand the complexity of mastering a discipline, but they can also relate their discipline (language) to the learning of most other disciplines. Thus, the language testing specialist can work well with teachers in other disciplines who are involved in test development projects in their discipline. This is particularly important if a test is to be translated or adapted. Appropriate testing policies must be developed and the language testing specialists can contribute readily to such policy development.

#### XIV References

- Alderman, D.L.** 1982: Language proficiency as a moderator variable in testing academic aptitude. *Journal of Educational Psychology* 74, 580–87.
- Alderson, J.C. and Urquhart, A.H.** 1983: The effect of student background discipline on comprehension: a pilot study. In Hughes, A. and Porter, D., editors, *Current developments in language testing*. London: Academic Press, 121–28.
- Alderson, J.C. and Urquhart, A.H.** 1985: The effect of students' academic discipline on their performance on ESP reading tests. *Language Testing* 2, 192–204.
- Bernard, H.R.** 1988: *Research methods in cultural anthropology*. Newbury Park, CA: Sage.
- 1994: *Research methods in anthropology: qualitative and quantitative approaches*. 2nd edition. Newbury Park, CA: Sage.
- Brislin, R.** 1970: Back translation for cross-cultural research. *Journal of Cross Cultural Psychology* 1, 185–216.
- 1976: *Translation: applications and research*. New York: Gardner.
- 1986: The wording and translation of research instruments. In Lonner, W.J. and Berry, J.W., editors, *Field methods in cross-cultural psychology*. Newbury Park, CA: Sage, 137–64.
- Brislin, R., Lonner, W.J. and Thorndike, R.L.** 1973: *Cross-cultural research methods*. New York: John Wiley.
- Cummins, J.** 1984: Wanted: a theoretical framework for relating language proficiency to academic proficiency among bilingual students. In Rivera, C., editor, *Language proficiency and academic achievement*. Clevedon, UK: Multilingual Matters, 2–19.

- 1  
2  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782
- Elder, C.** 1997: What does test bias have to do with fairness? *Language Testing* 14, 261–77.
- Garcia, T., del Rio Parent, L. et al.** 2000: *Study of a dual language test booklet in 8<sup>th</sup> grade mathematics: final report*. Washington, DC: American Institutes for Research.
- Gunnarsson, B.** 1978: A look at the content similarities between intelligence, achievement, personality, and language tests. In Oller, J. W., Jr. and Perkins, K., editors, *Language in education: testing the tests*. Rowley, MA: Newbury House, 17–35.
- Hale, G.** 1988: Student major field and test content: interactive effects on reading comprehension in the Test of English as a Foreign Language. *Language Testing* 5, 49–61.
- Hambleton, R.K.** 1993: Translating achievement tests for use in cross-national studies. *European Journal for Psychological Assessment* 9, 57–68.
- Joint Committee on Standards for Educational and Psychological Testing.** 1999: *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Liu, K.K., Anderson, M.E. and Swierzbis, B.** 1999: *Bilingual accommodations for limited English proficient students on statewide reading tests: phase I*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Marin, G. and Marin, B.V.** 1991: *Research with Hispanic populations*. Newbury Park, CA: Sage.
- Office of Elementary and Secondary Education.** 1996, April: *Title I, Part A policy guidance: Improving basic programs operated by local educational agencies, guidance on standards, assessments, and accountability*. Washington, DC: USA Department of Education.
- 2000, April: *Summary guidance on the inclusion requirement for Title I final assessments*. Washington, DC: USA Department of Education.
- 2002, July: Final regulations. *Federal Register* 67, Pages 45037–47.
- Oller, Jr., J.W. and Perkins, K.,** editors, 1978: *Language in education: testing the tests*. Rowley, MA: Newbury House.
- Rivera, C., Stansfield, C.W., Scialdone, L. and Sharkey, M.** 2000: *An analysis of state policies for the inclusion and accommodation of English Language Learners in state assessment programs during 1998–99*. Arlington, VA: George Washington University, Center for Equity and Excellence in Education.
- Sireci, S.G.** 1997: Problems and issues in linking assessment across languages. *Educational Measurement: Issues and Practice* 16, 12–19.
- Spolsky, B.** 1997: The ethics of gatekeeping tests: what have we learned in a hundred years? *Language Testing* 14, 242–48.
- Stansfield, C.W.** 1998: *A brief consideration of two test administration accommodations for LEP students: the use of one parallel column test booklet versus two different test booklets and the use of bilingual dictionaries versus glossaries*. Arlington, VA: ERIC Document Reproduction Service.

1 206 *Test translation and adaptation in public education*

- 2  
783 ——— 1997: *Experiences and issues related to the format of bilingual tests*.  
784 Springfield, VA: ERIC Document Reproduction Service, ED 423 306.  
785 **Stansfield, C.W.** and **Kahl, S.R.** 1998: Lessons learned from a tryout of  
786 Spanish and English versions of a state assessment. Paper presented  
787 at a symposium on multilingual versions of tests at the annual meeting  
788 of the American Educational Research Association. San Diego, CA.  
789 Springfield, VA: ERIC Document Reproduction Service, ED 423 306.  
790 **Stansfield, C.W., Bowles, M.** and **Rivera, C.** in progress: *Case studies of*  
791 *state policies and practices concerning test translation*. Washington,  
792 DC: George Washington University.  
793 **United States Congress.** 1994a: *Goals 2000: Educate America Act*. Public  
794 Law 103–227. Washington, DC: Government Printing Office.  
795 ——— 1994b: *Improving America’s Schools Act*. Public Law 103–382.  
796 Washington, DC: Government Printing Office.  
797 ——— 2002: *No Child Left Behind Act*. Public Law 107–110. (H.R.1)  
798 (Sections 1111(3)(C); 1111(b)(3)(F); and 1111(b)(5). Available  
799 online at: <http://www.ed.gov/legislation/ESEA02/107-110.pdf>  
800 **Werner, O.** and **Campbell, D.T.** 1970: Translating, working through  
801 interpreters, and the problem of decentering. In Naroll, R. and Cohen,  
802 R., editors, *A handbook of cultural anthropology*. New York:  
803 American Museum of Natural History, 398–420.  
804 **Woods Chapman, C.** 1996: Assessing language minority students for  
805 improved accountability. Paper presented at the annual meeting of the  
806 National Council on Measurement in Education. Chicago, IL.  
807