

**Translation Accommodations Framework for Testing English Language Learners in
Mathematics**

Guillermo Solano-Flores

University of Colorado at Boulder

Submitted to the Smarter Balanced Assessment Consortium (SBAC)

September 18, 2012

Table of Contents

Executive Summary	3
Scope of the Assessment Translation Framework	5
Targeting Languages for Test Translation	10
ELLs' Native Languages	10
American Sign Language	11
Understanding ELL Populations	14
Linguistic Heterogeneity of ELL Populations	14
Identification of English Language Learners	17
Identification of English Language Learners with Disabilities	18
The Nature of Test Translation	20
Testing as a Communication Process	20
Translation and Construct Equivalence	22
Translation and Dialect	24
Translation and Register	25
Translation as a Testing Accommodation	29
Types of Test Translation Accommodations	30
Evaluating Test Translation Accommodations	32
Safety of Untargeted Test Takers	33
Sensitivity to Individual Test Takers' Needs	33
Fidelity of Implementation	34
Usability	35
Ensuring Validity and Fairness of Translation Testing Accommodations	37
Assessment Translation Models	42
Translation Team	43
Translation Preparation Activities	46
Independent Translations and Translation Reconciliation	47
Translation Review/Revision	47
Cognitive Interviews	54
Format Design	55
Word Tagging	56
Final Version of the Translation	57
Follow-Up Activities	57
Systematic Development of Test Translation Accommodations	59
Assessment Translation Error Dimensions	59
Assessment Translation Specifications	63
Translation and Translation Support Materials	65
Documenting the Process of Test Translation	67
Final Remarks	70
Notes	71
References	72

Executive Summary

The present framework is developed under contract with the Smarter Balanced Assessment Consortium (SBAC) as a conceptual and methodological tool for guiding the reasonings and actions of contractors in charge of developing and providing test translation accommodations for English language learners.

The framework addresses important challenges in the development and use of effective translation accommodations for English language learners. Many of these challenges are directly related to the fact that translation is a complex activity. Other challenges stem from the fact that this complexity is often underestimated and from the fact that the process of test development and the process of test translation are often viewed as unrelated and are limited by tight timelines.

According to this framework, test translation is a complex endeavor that goes beyond the simple act of translating test materials and involves more professionals than those in charge of translating tests. According to the framework, successful test translation projects take into consideration factors such as the tremendous linguistic heterogeneity of populations of English language learners, the potential fallibility of translation accommodations, and the need to coordinate efforts with agencies and colleagues who are external to the process of test translation yet whose actions influence the integrity of test translations.

Four translation accommodations are identified as viable in the testing of English language learners: Test Version in the Native Language, Side-by-Side Bilingual Version of the Test, Directions Translated into Native Language, and Bilingual Glossary. Their limitations and possibilities are discussed in terms of four validity and fairness dimensions: Safety of Untargeted Test Takers, Sensitivity to Individual Takers' Needs, Fidelity of Implementation, and Usability.

A basic translation model is offered for each of these translation accommodations. The four models rely heavily on the use of multidisciplinary teams, the use of cognitive interviews on samples of translated items, and the focus on error as critical to evaluating and refining test translation.

The document discusses the nature of translation support materials that should be made available to professionals participating in test translation projects and the need for translation specifications documents that specify the lexical and discursive characteristics of the translated materials. The former ensure that the process of test translation is informed by knowledge of the standards, skills, and knowledge assessed by the translated items; the latter ensures standardization in the characteristics of test translation—an important aspect to address in massive translation projects in which different sets of professionals translate and review the translations of different sets of items.

The framework provides a list of the documents and pieces of evidence that, in addition to the translated materials, should be provided to document the process of development of test translation accommodations.

Scope of the Assessment Translation Framework

The Smarter Balanced Assessment Consortium (SBAC) is in charge of developing assessments aligned to the English language arts/literacy and mathematics Common Core Standards in Grades 3 to 8 and high school. According to the consortium's timeline, the implementation of the assessment system will start in the 2014-2015 school year.

One of the stated commitments of the consortium is the fair and valid testing of two special populations, students with disabilities and English language learners (ELLs)—students who are still developing English as a second language while they continue developing their first, native language.

SBAC intends to ensure the accessibility of test items (also called, tasks in this document) to these students through the use of testing accommodations. Testing accommodations can be defined as changes to the ways in which tests are administered to ensure that students with special needs gain access to the content of assessments—to ensure that these students are able to understand what assessment tasks or items ask them so that they can demonstrate their knowledge. Testing accommodations should not alter the constructs measured by the assessments, should not lead the students in their responses (e.g., by giving away the correct answers), and should not give an unfair advantage to the students who receive the accommodations over students who do not receive the accommodations.

ELLs are the focus of this framework. Test translation accommodations are among the testing accommodations to be used with ELL students in the SBAC mathematics assessments. Test translation accommodations are intended to address limited proficiency in English as a condition that may unfairly affect the students' understanding of the items, adversely affect their performance on tests, and threaten the validity of the measures of their achievement.

Making translation accommodations available as a testing accommodation for English language learners poses multiple, formidable challenges. Some of them are practical, others methodological. First, because translations need to be made on final versions of documents, the timelines for developing translations may be restricted by the timeliness with which the final versions of items in English are available. A tight timeline for test translation may limit the opportunity for proper and extensive review, thus affecting the quality and validity of the translated instruments.

Second, because SBAC is to generate thousands of items, they are likely to be translated by multiple teams of translators. Without an appropriate set of actions for selecting and training translators and without a good set of translation procedures, the quality and style of the translations may be seriously compromised. This may constitute another important threat to the validity of the instruments.

Third, because ELLs vary tremendously in their reading and writing proficiencies in English and have multiple schooling histories in English, many of them may not benefit at all from translation accommodations. Classifications of students according to a few levels of English proficiency are not sensitive to the ability of ELLs to read and write in English, especially in the context of academic English. Unduly assigning this form of accommodation to ELLs may be more harmful than beneficial.

This assessment translation framework is intended to provide SBAC decision makers and contractors with the reasonings and procedures needed to make appropriate decisions about translation accommodations and to properly implement translation as a valid form of testing accommodation for ELL students. The framework's theoretical stand can be characterized as systemic, critical, multidisciplinary, and bottom-up.

First, the assessment framework's perspective is systemic because it views test translation as a process that involves multiple actors in the assessment system, not only the translators. In the context of SBAC and the testing of ELLs, translation is not only about creating another language version of a test. Translation also involves making sound decisions about the specific ELLs who are to be given test translation accommodations. It also entails meticulous coordination work with various SBAC system components not directly involved in the translation process. This coordination is intended to secure conditions for effective translation, including that the original English versions of the assessments are made available to test translators in a timely manner and that the formats used in the tasks in English are designed with consideration for the format requirements of other languages.

Second, the assessment framework's perspective is critical because it recognizes the multiple factors that shape the effectiveness of testing accommodations. Among these factors are the linguistic heterogeneity of ELLs (their multiple patterns of proficiency in English and in their native language)—which limits the effectiveness of blanket approaches consisting of giving the same accommodation to all ELL students—and the fact that many accommodations are likely to be poorly implemented.

Third, the assessment framework's perspective is multidisciplinary because it is based on knowledge from different professional fields, mainly psychometrics, translation, second language acquisition, and sociolinguistics. Professionals from these fields have different yet mutually complementary views of language and bilingual individuals. Addressing and integrating these multiple perspectives is critical to producing translations that address, among other things, the intrinsic relation between language, academic language, and the target constructs; the formal properties of test translation; the multiple patterns in which ELLs, as

emergent bilinguals, develop their first and second languages; and the ways in which the characteristics of multiple linguistic groups are shaped by issues such as language contact and language variation.

Finally, the assessment framework's perspective is bottom-up because it intends to be sensitive to dialect variation within ELL students' first languages. The purpose of translating tests for ELLs is to eliminate limited proficiency in English as a threat to the validity of academic achievement measures. Therefore, a translation that reflects exclusively the style of the translator and ignores language usage by the communities of users of a language may fail to serve the goal of ensuring accessibility. Available evidence from research in which ELLs are tested across dialects indicates that the performance of these students is as sensitive to first language dialect differences as it is to language (native and second language) differences (Solano-Flores & Li, 2006). Thus, even in cases in which it is appropriate to test ELLs in their first language, test translation needs to be sensitive to dialect variation. The old approach in which one translator and one translation reviewer work in isolation to the best of their ability may be less expensive, but is not sensitive to dialect variation in the target language (Harkness & Schoua-Glusberg, 1998), even when the translators are highly qualified. Thus, consistent with modern test translation practice, the framework promotes the use of multidisciplinary teams of professionals who review the translators' translations and, depending on the form of translation accommodation, may also participate in different stages of the translation process.

The assessment translation framework incorporates major and (relatively) recent conceptual and procedural developments in the field of test translation. Many of these developments are part of the procedures currently used in large-scale testing projects, as is the case of PISA—the Programme of International Student Assessment funded by the Organization for Economic

Cooperation and Development. In addition to an emphasis on the use of multidisciplinary teams (see above), the framework does not support the use of back translation, a procedure for test translation verification that is now discredited. In the back translation procedure, the translated version of the test is translated back to the source language and the first and back-translation versions of the source language are compared to ensure that the content of the text has been preserved. The limitations of this procedure are well documented. Experience from international test comparisons has shown that back translation may recover the original text without detecting translation error (Grisay, 2003).

It is customary to distinguish between translation and interpretation, respectively, to refer to the textual (printed/written) and oral (spoken) forms of language in which communication takes place. However, for the sake of simplicity, this document treats interpretation as a form of translation. This makes it possible to discuss with ease the variety of testing accommodations involving translation, some of which take place both across languages and across the textual and oral forms of language.

For the sake of simplicity, *test* and *assessment* and *testing* and *assessment* are used in this framework as pairs of interchangeable terms.

Targeting Languages for Test Translation

Whereas deciding which languages should be selected for translation accommodations is not within its scope, the framework provides some reasonings for making sound decisions about the languages that should be targeted. In general, these decisions should be based on both the numbers of students that translation accommodations are likely to serve and the appropriateness of providing these accommodations in those languages. Two cases are considered: selecting ELLs' native languages and developing translation accommodations for users of American Sign Language and Signed Exact English

ELLs' Native Languages

About 10% of the students in Pre-K and K-12 enrolled in public schools in the U.S. have limited proficiency in English (Kindler, 2002). Of the dozens of languages spoken at home by ELLs in the U.S., only a few can be selected for translation accommodations due to limited human and financial resources. Necessarily, Spanish is the top priority as a target language for translation accommodations. Over 73% (more than 3.6 million) of the ELL students in the U.S. are users of Spanish as a native language (Migration Policy Institute, 2010).

Following Spanish, the most frequently spoken languages used by ELLs are Chinese, Vietnamese, and Haitian-Creole (respectively, 3.8%, 2.7%, and 2.1% of the ELLs). None of the dozens of languages spoken by ELLs in the U.S. account for more than 2% of the total number of ELLs in the U.S. However, these percentages are national and they are not consistent across states. Thus, while Spanish is the most frequent language among ELLs in 43 states, Ojibwa, Somali, Dakota, Bosnian, Ilokano, Yupik, and a mix of American Indian languages are the languages most spoken by ELLs respectively in North Dakota, Maine, South Dakota, Vermont, Hawaii, Alaska, and Montana (Migration Policy Institute, 2010). These proportions reflect both

long-term and recent historical and demographic trends. The resulting diversity poses a serious challenge to state consortia, as each consortium state may have a unique set of potential ELL native languages to consider.

To deal with these challenges, three criteria can be used in combination to determine which ELL native languages should be targeted for translation accommodations. The first (and obvious) criterion is frequency. Deciding what languages, in addition to Spanish, should be served by translation accommodations should be based on careful consideration of the percentages of users of ELL native languages both by state and across consortia states. The second criterion is feasibility. It is not appropriate to attempt to generate translation accommodations for languages for which highly qualified translators are difficult to recruit. The third criterion is the stability of the population of speakers of the native language. Translating tests targeted to highly mobile groups may not actually result in obtaining more or better data on their academic achievement.

American Sign Language

Most of the issues discussed in this framework are applicable to translating SBAC assessments into American sign language (ASL), which has the properties of any other language. In principle, it is possible to administer SBAC in ASL via an interpreter who translates the content of items into ASL and who may also record in writing the students' responses given in ASL. However, in order to make sound assignment decisions concerning this form of testing accommodation, it is important to keep in mind that the role ASL plays in ensuring access to instruction for deaf/hard of hearing students is different from its potential role in ensuring these students' access to the content of items in an assessment.

In the mainstream classroom context, a great deal of instruction takes place through social interaction involving the listening and speaking modes of English, mainly in the form of

teachers' unscripted verbalizations and teacher-student and student-student conversations. In schools and programs for deaf and hard of hearing students, the classroom context is similar, with one exception: ASL is predominantly the language of instruction and social interactions. In a mainstream/public school classroom context, an ASL interpreter is used to provide deaf/hard of hearing students' access to those unscripted verbalizations and participation in those conversations. However, English is the language of instruction. Thus, this use of ASL does not prevent these students from developing and using English in the reading and writing modes (e.g., through reading and written assignments and classroom-based, paper-and-pencil, or computer-administered tests). In contrast, in the context of assessment, the content of items is scripted. Accessibility is shaped by the proficiency of students in English in the reading mode and in English in the writing mode—respectively, their ability to understand printed text in English and to give their answers in written English. It should not be assumed that deaf/hard of hearing students are limited in their proficiency in English in the reading and writing modes. Moreover, the fact that they receive ASL support during instruction does not necessarily mean that they are tested more fairly in ASL than in English. There are numerous challenges with access to content and communication posed by use of an interpreter and attempts to provide a signing accommodation to deaf/hard of hearing students. Testing a deaf/hard of hearing student in ASL is justified only when three conditions are met: 1) the student has a history of instruction in ASL; 2) the student's proficiency in English in the reading mode is limited and lower than his/her proficiency in the receptive (viewing) mode of ASL; and 3) the student's proficiency in English in the productive (writing) mode is limited and lower than his/her proficiency in the productive (signing) mode of ASL.

Another issue to be considered in the testing of deaf/hard of hearing students in ASL concerns the validity of interpretations of their scores. While printed text ensures standardization in testing, the administration of a test in ASL is likely to vary considerably across students due to multiple factors such as the skills of the ASL interpreters or their familiarity with academic language or the topic assessed. Thus, decisions on the testing of deaf/hard of hearing students through ASL translations should be based first, on information concerning their strengths and weaknesses in the reading and writing modes in English and in the receptive and productive modes in ASL; and second, on information on the qualifications and experience of the individuals that are to serve as interpreters.

Understanding ELL Populations

Properly understanding the characteristics of ELL populations is key to making sound decisions about the adequacy of test translation accommodations. Unduly assigning this form of accommodation to ELLs may be more harmful than beneficial. Three aspects of the complexity of ELL populations deserve consideration: the linguistic heterogeneity of ELL populations, the challenges of accurately identifying ELL students, and the challenges of identifying ELL students with disabilities. (Note 1)

Linguistic Heterogeneity of ELL Populations

The linguagrams shown in Figure 1 are intended to help one to understand ELLs and the heterogeneity of ELL populations. Linguagrams are conceptual tools consisting of symmetric bar graphs that compare the proficiency of individuals in their first and second language for each of the four language modes: listening, speaking, reading, and writing (Solano-Flores & Gustafson, 2013). Several cases are shown.

The figure helps one to appreciate that the term, *English language learner* may be misleading, as it may evoke Case A—an individual who is not proficient at all in English and is fully proficient in his first language. In reality, Case A would be that of an individual who has completed the development of his first language, has been educated in that language, has never had any exposure to English, and lives in a society in which the predominant language is his first language.

In reality, ELLs can be characterized more accurately as *emergent bilinguals* (García & Kleifgen, 2010). That term conveys the fact that they are developing English as a second language while they continue developing their first language.

Thinking about ELLs as bilingual individuals may be difficult for some because *bilingual* is a term that may evoke Case B, an individual who is fully proficient in the four language modes in her two languages. In reality, the cases of bilingual individuals who are equally proficient in two languages are rare (see Mackey, 1962; Butler & Hakuta, 2006). The term, *proficiency* may be wrongly interpreted as implying that an individual's level of proficiency in a language is the same across the four modes of that language: listening, speaking, reading, and writing. In reality, bilingual individuals are virtually always unequally proficient across language modes in both their first and their second language (Grosjean, 1985; Valdés & Figueroa, 1994).

Cases C-F are more realistic examples of bilingual individuals. Case C is hardly representative of the ELLs in the U.S. This could be an individual who has grown up and lives in a society in which the predominant language is his or her first language and has studied English as a foreign language while living in that society, mainly through reading and writing, with limited opportunities for developing conversational skills.

In contrast, Cases D, E, and F are more realistic examples of ELLs in the U.S.—individuals who are developing both English and their first language, as pointed out before. Different social contexts, demographic factors, kinds of exposure to each language, and schooling histories shape their proficiency in each linguistic mode in each language, thus producing multiple patterns of bilingualism among ELLs.

Whereas limited proficiency in a second language is sometimes wrongly confused with deficiency, evidence on bilingual development does not support this view. For example, in terms of vocabulary development, young bilinguals are able to recognize or speak, in the two languages combined, the same number of words or even more words than monolinguals of their same age in one language (Oller, Pearson, & Cobo-Lewis, 2007). That is the reason that the bars

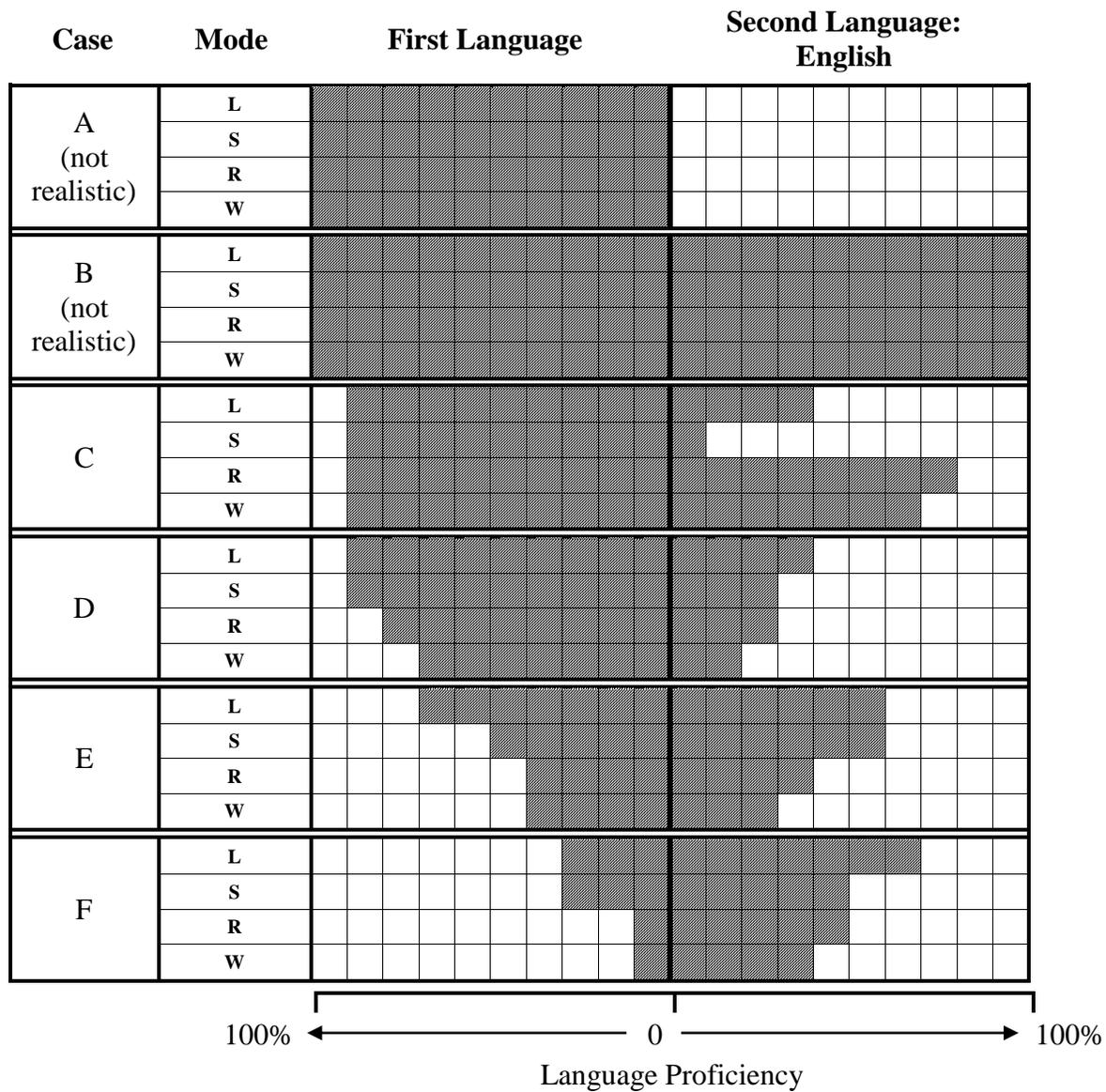


Figure 1. Language proficiency patterns of one monolingual and five bilingual individuals.

for the listening and speaking modes for Cases D, E, and F add up to 100 or more than 100 when the first language and the second language are taken together. Often, measures of English proficiency are wrongly used to make inferences about students' overall language development, resulting in a partial and inaccurate picture of the linguistic capabilities of these students.

Identification of English Language Learners

The tremendous heterogeneity of ELL populations illustrated by Figure 1 explains why categories used to describe English proficiency (e.g., “limited English proficient,” “fluent English proficient”) cannot describe each student’s specific set of strengths and weaknesses in English. Measures of English proficiency and official definitions of English language learners (e.g., NCLB, 2001) are associated with attempts to meet legal requirements to serve certain segments of the student population (Durán, 2008) but are limited in their effectiveness to provide a detailed picture of each ELL student’s English proficiency (see Abedi, 2004, 2007b).

Figure 2 represents the implications of these limitations. The confusion matrix shows whether classifications of students as either ELL or non-ELL based on a test of English proficiency (columns) are consistent with the students’ actual condition as either ELL or non-ELL (rows) that would be possible to determine if a sufficient number of measures of English proficiency were available (see Solano-Flores & Gustafson, 2013).

		Classification Based on a Test of English Proficiency	
		ELL	Non-ELL
Actual Condition	ELL	accurate	false positive
	Non-ELL	false negative	accurate

Figure 2. Confusion matrix representing classifications of students as ELLs or non-ELLs.

Accurate classifications occur when students who should be classified as ELLs are classified as ELLs, or when students who should be classified as non-ELLs are classified as non-ELLs. False positive classifications occur when students who should not be classified as ELLs are classified as ELLs. False negative classifications occur when students who should be classified

as ELLs are not classified as ELLs. But, as we have seen, these crude classifications mask a world of difference from student to student within the two categories.

Cases of false positive classifications are of special interest for this assessment translation framework. While the act of translating tests is not directly related to the act of classifying students according to their English proficiency, certain translation accommodations may be detrimental rather than beneficial for students who are wrongly classified as ELLs.

Identification of English Language Learners with Disabilities

An important sub-group of ELLs is that of ELLs with disabilities. The performance of ELL students with disabilities on large-scale tests is lower than that of non-ELL students and ELL students without disabilities (Abedi, 2006). Certain guidelines exist for providing accommodations to students with disabilities (e.g., Thurlow, Elliot, & Ysseldyke, 2003). For ELL students with disabilities, accommodations intended to address their disabilities should be used in combination with accommodations intended to address language proficiency. In practice, however, this ideal is difficult to meet because many accommodations are limited in their effectiveness to serve the needs of students with disabilities, and the fidelity with which they are implemented varies considerably (Sireci, Scarpati, & Li, 2005).

Important challenges in the identification of students with disabilities shape the effectiveness of both the testing of ELL students with disabilities and the testing of ELL students without disabilities (see Abedi, 2007a). These challenges stem from the complexity of accurately classifying students according to both English proficiency and the presence or absence of disabilities. As the confusion matrix in Figure 3 shows, multiple forms of misclassification are possible. Concerns about false positive classifications (lower diagonal of the matrix), especially ELL students wrongly classified as students with disabilities, result from consistent evidence that

ELL students are overrepresented in special education programs. This overrepresentation is an effect of multiple factors, including the lack of adequate schooling options for these students, cultural misunderstanding, confusing limited English proficiency with learning disabilities, and misinformed special referral decisions made by educators (Artiles & Ortiz, 2002; Harry & Klingner, 2006; Klingner & Solano-Flores, 2007).

Classification Based on Few or Limited Sources

		ELL		Non-ELL	
		With disabilities	Without disabilities	With disabilities	Without disabilities
		Actual Condition	ELL	accurate	false negative
Non-ELL	false positive		accurate	false negative	false negative
Actual Condition	ELL	false positive	false positive	accurate	false negative
	Non-ELL	false positive	false positive	false positive	accurate

Figure 3. Confusion matrix representing classifications of students as ELLs with and without disabilities or non-ELLs with and without disabilities.

The Nature of Test Translation

Translation can be defined as the activity intended to convey meaning in a language other than the language in which meaning was originally expressed. It entails decoding meaning in one language—the source language—and recoding it in another language—the target language. Languages encode the experiences of their users over time, and culture shapes the ways in which their users use language. Consequently, effective *decoding* and *recoding* entails not only using the vocabulary, grammar, and other conventions of the source and target languages but also understanding the ways in which culture shapes meaning in each language. For example, different sets of cultural experiences may influence the ways in which different users interpret the same piece of translated text. In the context of testing, decoding and recoding meaning entails, in addition, understanding the construct or constructs measured by a given test and the way in which its linguistic characteristics influence its difficulty.

Testing as a Communication Process

To understand the complexities of test translation, it is appropriate to view assessment as a communication process. Assessment can be viewed as a process in which the test poses questions, and students answer them. The test functions as an interlocutor in a more direct way than other kinds of texts that students encounter. Students' answers to those questions are interpreted to make inferences about the students' academic achievement (Solano-Flores, 2008). In the classroom context, the teacher asks questions and interprets students' responses. Even when this interaction takes place through formal, paper-and-pencil tests in the classroom, this process parallels an interaction through conversation. In this interaction, how questions are asked to students and how students' responses are interpreted is influenced or informed by the teacher's knowledge of the instructional context, which comprises the students' strengths and weaknesses

in the language of instruction, the content that has been taught, the way in which this content been taught, and many other pieces of information.

In large-scale assessment, the communication process does not take place as a one-on-one interaction. The individuals who establish the content to be assessed, write the items of a test, and score students' responses are not the same. Also, to ensure standardization, the ways in which students are asked questions and the ways in which students' responses to those questions are interpreted are the same for all individuals in the population tested and cannot reflect the variety of instructional contexts in which instruction takes place.

When translation accommodations are used to test ELLs in large-scale assessment, this communication process is even more complex and involves even more actors. The individuals who write the test items in English, translate those items, interpret students' responses, and decide which students should be provided with translation accommodations are not the same. In addition, the ways in which students are asked questions and the ways in which students' responses to those questions are interpreted are limited not only in their sensitivity to the variety of instructional contexts in which instruction takes place but also in their sensitivity to the variety of students' linguistic backgrounds and the variety of linguistic contexts in which they learn. Given this complexity, in large-scale assessment, ELL students who are given test translation accommodations will be better served if test translation is viewed as a process that involves multiple professionals, rather than simply the tasks performed by translators. Thus, effective use of translation accommodations involves not only good translators but also effective coordination between translators and other professionals involved in the process of testing. Table 1 examines the multiple issues of test translation in ELL testing in terms of the question, *Who is given tests in what language, by whom, when, and where?* (Solano-Flores, 2008). Failure to implement or

poor implementation of any of its components affects the integrity of the entire process of test translation and, consequently, the validity of the measures of academic achievement for ELLs (Solano-Flores, 2009).

Table 1

Translation Issues in the Process Of ELL Testing, Defined by the Question, Who Is Given Tests in What Language, by Whom, When, and Where?

Assessment process component	Issues related to the process of test translation
<i>Who...</i>	<ul style="list-style-type: none"> • Procedures and criteria used to identify ELLs • Criteria for determining cases in which ELLs are to be tested with translation accommodations • Linguistic groups for which translation accommodations will be made available
<i>...is given tests...</i>	<ul style="list-style-type: none"> • Test translation procedures • Translation accommodations to be used
<i>in what language,</i>	<ul style="list-style-type: none"> • Languages into which tests are to be translated • Approaches to addressing language variation due to dialect • Approaches to addressing academic language
<i>by whom,</i>	<ul style="list-style-type: none"> • Qualifications of individuals who translate tests • Qualifications of individuals who review test translations • Qualifications of individuals who provide translation accommodations • Qualifications of individuals who score the responses of students who are given translation accommodations
<i>when,</i>	<ul style="list-style-type: none"> • Time at the process of development of English as a second language development at which ELLs are tested • ELL students' histories of schooling in English • ELL students' histories of schooling in the first language
<i>and where?</i>	<ul style="list-style-type: none"> • Conditions needed to properly use translation accommodations with ELLs

Translation and Construct Equivalence

Language is the system of human communication through which combinations of sounds (or hand movements, in the case of sign language) are used to represent meaning. Languages encode experience and meaning according to a set of arbitrary rules and conventions. *System* refers to the fact that the features of a language are organized and related to each other. It is because

languages are systems that it is possible to express the same idea in multiple (although not exactly the same) ways in the same language (e.g., *Rose was hired!* and *Rose got the job!*). *Encoding experience and meaning* refers to the fact that different languages evolve in different ways according to the communication needs of their users. These communication needs are determined by environmental and societal characteristics. *Arbitrary* refers to the fact that the sounds or symbols and the rules for combining those symbols to communicate do not have a direct relationship to the kinds of experience or meaning encoded. For instance, the sounds of the words “apple” or “manzana” have nothing to do with the characteristics of apples. Words, rules for verb conjugations, or pronunciation, or any other features of different languages are equally arbitrary. No language is more “logical” than others.

Because languages encode experience and meaning in different ways (Greenfield, 1997; Nettle & Romaine, 2002), many ideas are not expressed with the same level of precision in different languages. Word games, jokes, and riddles are a good resource to illustrate this:

Why are there fences around cemeteries?... Because people are dying to get in.

While *dying* is used to refer to *utterly wanting* in many languages, the use of this idiomatic expression is not as frequent as it is in English. The translation of the joke in another language may be understood, but does not convey exactly the same meaning or have the same effect as a joke (in some other languages, it might not be understood at all). Something is lost or changed in translation.

The fact that languages do not encode experience and meaning in the same ways is the main challenge faced in any assessment translation endeavor. As a result of translation, the same item may end up measuring different knowledge or skills in different languages (American Educational Research Association, 1999). Effective translation maximizes the likelihood that

students tested in different languages attach the same meaning to the target construct—the specific skill, knowledge, or ability an item is intended to measure (see Ercikan, 2002; van de Vijver & Poortinga, 2005).

In addition to the fact that languages do not encode experience and meaning in the same ways, assessment translators face the challenge that culture shapes the ways in which speakers of the same language use it. Two aspects of this language variation need to be discussed, dialect and register.

Translation and Dialect

While *dialect* is commonly used in a derogatory form to refer to some “corrupt” version of a language, the term actually refers to any variety of a language distinguishable from other varieties by features such as pronunciation, vocabulary, grammar, and discursive forms and by the frequency of these features (see Wolfram, Adger, & Christian, 1999). The variety of English used by the royal family in England—a highly prestigious variety of English—and the variety of English used by miners in Appalachia are two dialects among the many English dialects in the world.

The term, *standard* (as in *standard English*) is used to refer to a prestigious dialect of a language. Linguistically, a standard dialect of a language is not any better than other dialects and is not necessarily used or understood equally by all the speakers of that language. Regardless of social status, any dialect is a complex and sophisticated rule-governed system of communication (Crystal, 1997). These notions are important for test translators to take into consideration, so that they do not unduly favor a given dialect of the target language over others, based on the wrong premise that such dialect is understood by all students in the target linguistic group.

A case in point is the Spanish translation of tests. The Iberian Spanish dialect (i.e., the Castilian Spanish used in Spain) may be wrongly assumed to be the “correct” dialect to use in Spanish translations of tests given to native Spanish-speaking ELL students in the U.S. However, the overwhelming majority of native Spanish speaking ELLs in the U.S. do not use this dialect. Many idiomatic expressions, colloquialisms, words, discursive forms, and even tenses and conjugation forms in Iberian Spanish are unfamiliar to them. Giving these students an Iberian version of Spanish of their tests may fail to support students to gain access to the content of items. These considerations are supported by evidence from research in which ELLs are tested across dialects. As mentioned, the performance of ELLs students is as sensitive to first language dialect differences (i.e., standard and local dialects) as it is to language (native and second language) differences (Solano-Flores & Li, 2006).

Because of the vastness of language, attempts to characterize dialect through lexical analyses may prove to be time consuming and costly in the context of test translation. Thus, rather than attempting to identify *the dialects* of a language into which a test should be translated, it is more practical to be sensitive to *dialect issues* by involving the users of different dialects of the target language in the process of test translation.

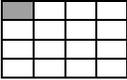
Translation and Register

Register is the variety of a language and set of forms of representation used in a specific context, as is the case of the context of mathematics. While it tends to be precise, due to the fact that it originates from specialization of human activity (Halliday, 1978), the characteristics of a register vary within the same language. As Table 2 shows, subtle but very important variations in notation and usage may shape the ways in which ideas are represented in the same language, as

is the case of the comma, used in some Spanish speaking communities as a decimal point (see Solano-Flores, 2011).

Table 2

Forms of Representation of a Fraction in English and in Spanish

Language	Form of representation			
	Quotient	Decimal	Graphic	Verbal
English	$1/16$	0.0625		<i>one sixteenth</i>
Spanish	$1/16$	0.0625 or 0,0625	(not common)	<i>un décimosexto</i> (formal) or <i>un dieciseisavo</i> (informal)

Because of dialect variation and differences in register across cultures within the same language, assessment translation projects need to take into consideration the characteristics of the target linguistic groups. The selection of translators; the development of a set of translation specifications; the use of multidisciplinary, multicultural teams of test translation and test translation review/revision teams; and the use of consensus-based procedures—all topics discussed later in this framework—are critical to addressing this linguistic diversity and producing optimal test translations.

On certain occasions, issues of language variation in test translation for ELLs are resolved by using the dialect and varieties of register used in what translators believe are the ELL students' countries of origin. While well intentioned, the strategy may be limited in its effectiveness to provide ELLs with the intended linguistic support. The reason is that, contrary to commonly-held beliefs, the majority of ELLs (76% in elementary school and 56% in middle schools) are U.S.

native-born (Capps et al., 2005); and over 50% of ELLs in public secondary school are second- or third-generation U.S. citizens (NEA, 2008). Given these demographics, the characteristics of ELLs' native languages are shaped by the American culture and by contact with English. Thus, the translation of a test should be sensitive to the characteristics of the ELLs' native languages, as they are used in the U.S., not in other countries.

Potentially, the challenges resulting from language variation due to dialect and to multiple varieties of register can be addressed through *localization*. Originated in the context of marketing and propelled by globalization, the term refers to the adaptation of a product according to the characteristics of a given region. In the context of assessment translation, localization refers to the process of adapting the linguistic features of test items to the ways in which language is used in a specific context, such as a school or a school district. To localize a test, with facilitation from project staff, educators from a given school or school district discuss the characteristics of the items and make consensus-based decisions on the ways in which the linguistic features of the items need to be modified, so that they reflect the characteristics of the dialect and register used in their community.

While there is evidence that localization can be an effective form of testing accommodation for ELLs (Solano-Flores, Li, Speroni, Rodriguez, Basterra, & Dovholuk, 2007), this evidence is limited to localization in English, not translated tests. In addition, while promising, localization poses various challenges. One challenge is that the process of localization is time consuming. Another challenge is that, as with translation, the constructs that items are intended to measure may be altered when their linguistic features are modified. Close supervision from project staff during the process of localization is needed with the intent to ensure construct equivalence across the original and localized versions of the test. Finally, the procedure requires that test secure

material be shared among educators from multiple communities, a circumstance that may pose serious logistical challenges to ensuring test security.

Translation as a Testing Accommodation

Important experiences with test translation have originated from international test comparisons such as the Programme for International Student Assessment (PISA) and Trends in Mathematics and Science Study (TIMSS). In these international comparisons, test items are developed in English and/or another language, then they are translated into the language or languages of each participating country. As a result of these experiences, procedures for translating tests have been refined and continue to evolve (see Hambleton, 2005). For example, as mentioned before, back translation (a translation verification procedure in which the translated version of the instrument is translated back to the original language and the original and back-translation versions are compared to determine if the content has been preserved) is no longer regarded as a procedure that can be used to warrant construct equivalence (see Grisay, 2003).

Many lessons about test translation can be learned from PISA, TIMSS, and other international test comparisons for the purpose of translating tests for ELLs. However, it is important to keep in mind that the forms of linguistic diversity involved in the context of international test comparisons and in the context of testing ELLs in the U.S. are different. First, the native language of students participating in international test comparisons is the predominant language in the society in which they live and the language used in their schools. In contrast, the native language of ELL students in the U.S. is not the predominant language in the society in which they live and, typically, their native language is not used in their schools. Second, in the context of international test comparisons, the need for translation results from the need to test diverse linguistic groups in the language in which they receive instruction (and which is their first language in the majority of the cases). In contrast, in the context of ELL testing in the U.S., the need for translation results from lack of proficiency in the country's predominant language

(see Stansfield, 2003). Thus, in the context of ELL testing in the U.S., test translation is intended as a form of testing accommodation.

Types of Test Translation Accommodations

The term, *testing accommodations*, is used to refer to changes in a test or in the conditions in which tests are administered with the purpose of supporting ELL students (or others) to gain access to the content of tests, without giving them an unfair advantage over students who do not receive the accommodation and without altering the constructs measured (Kopriva, Emick, Hipolito-Delgado, & Cameron, 2007; Thurlow, 2007; Young & King, 2008).

The majority of state assessment systems authorize the use of accommodations for ELLs. Altogether, there are over 70 forms of testing accommodations that are used with these students and which can be classified under two broad categories: direct linguistic support and indirect linguistic support (Rivera, Collum, Shafer, & Sia, 2006). *Direct linguistic support accommodations* target the linguistic features of test items and can be delivered in the student's first language (as is the case of translations) or in English (as is the case of linguistic simplification). *Indirect linguistic support accommodations* target testing conditions that do not have to do with language (such as the time in which the test is administered or the seat assigned to the student to take the test) but which may contribute to supporting the student to better handle the linguistic load imposed by tests. Unfortunately, many forms of accommodations used with ELL students are questionable; their use is not supported by any evidence on their effectiveness and is borrowed from the field of special education (see Abedi, Lord, Hofstetter, & Baker, 2001; Rivera, Collum, Schafer, & Sia, 2006).

Table 3 provides a list of accommodations identified by Francis, Rivera, Lesaux, Kieffer, & Rivera (2006) as accommodations whose effectiveness has been documented. *Documented*

effectiveness means that an accommodation actually contributes to reducing the score gap between ELL and non-ELL students that is attributed to the ELL’s limited proficiency in English.

Table 3
Effective Accommodations for English Language Learners by Form of Linguistic Support.
Adapted from Francis, Rivera, Lesaux, Kieffer, & Rivera (2006).

Indirect Linguistic Support	Direct Linguistic Support	
	In English	Translation Accommodations
<ul style="list-style-type: none"> • Extended time allowed for test completion • Breaks offered between sessions 	<ul style="list-style-type: none"> • English glossaries • English dictionaries • Directions read in English • Linguistic simplification in English • Dictation of answers or use of a scribe 	<ul style="list-style-type: none"> • Test version in the native language • Side-by-side bilingual version of the test • Directions (written) translated into native language • Bilingual glossary <ul style="list-style-type: none"> ○ Printed glossary ○ Pop-up glossary ○ Audio glossary • Test taker responses in native language • Directions read in the student’s native language

For the purposes of this framework, the accommodations are grouped by type of linguistic support (direct or indirect) and, within the category “direct linguistic support,” as those provided in English and in the students’ first language. Notice that the majority of the accommodations listed in Table 3 take the form of direct linguistic support. Notice also that within the category “direct linguistic support” all the accommodations provided in the ELL’s first language involve translation or the use of interpreters.

The accommodations listed should be interpreted as families of forms of testing accommodations. For example, while it is easy to agree that a bilingual glossary provides translations (not definitions) of words, the accommodation may be implemented in multiple ways. (Note 2) What criteria will be used to determine the words to be included or excluded in the glossary? Will the glossary be available to students in printed form or electronically? What design characteristics will the document have? A test translation project should provide a detailed definition of the type or types of translation accommodations to be used and the methods used to create them.

Evaluating Test Translation Accommodations

Each type of translation accommodation has a unique set of advantages and disadvantages. Table 4 compares the six translation accommodations as to their likelihood to comply with four validity and fairness dimensions.

Table 4

Translation Accommodations and Their Likely Ability to Meet Validity and Fairness Dimensions

Translation Accommodation	Validity and Fairness Dimensions			
	Safety of Untargeted Test Takers	Sensitivity to Individual Test Takers' Needs	Fidelity of implementation	Usability
Test version in the native language	Low	Low	High	High
Side-by-side bilingual version of the test	High	High	Medium	Medium
Directions translated into native language	Low	Low	High	High
Bilingual glossary	High	High	High	Medium
Test taker responses in native language	Low	Medium	Low	Low
Directions read in the student's native language	Low	Medium	Low	Low

Safety of Untargeted Test Takers. This dimension refers to how likely a translation accommodation is to be safe for ELLs who do not need it but receive it. The terms “safe” and “safety” are used because ELLs who do not need an accommodation may actually be harmed by it; incorrect judgments may be made about their knowledge or skills on the basis of their response to an accommodation. The relevance of this dimension becomes apparent when we take into consideration that each ELL student has a unique pattern of strengths and weaknesses in both English and their first language and many students are wrongly classified as ELLs and their classifications vary across states.

In Table 4, Side-by-Side Bilingual Version of the Test and Bilingual Glossary are rated as being highly safe for untargeted test takers because they do not do any harm to students who are wrongly assigned these translation accommodations. In contrast, Test Version in the Native Language, Directions Translated into Native Language, Test Taker Responses in Native Language, and Directions Read in the Student’s Native Language are rated as having a low safety level for untargeted test takers. The performance of ELLs wrongly assigned to these translation accommodations might be lower than if they are not assigned any accommodation.

Sensitivity to Individual Test Takers’ Needs. This dimension refers to how likely a translation accommodation is to be sensitive to the specific set of linguistic needs of each ELL. A test taker-sensitive translation accommodation is not imposed; rather it is made available for the test taker to use optionally. The test taker can use the accommodation (or not) in ways that meet his or her specific needs.

In Table 4, Side-by-Side Bilingual Version of the Test is rated as being highly sensitive to individual test takers’ needs. Depending on the challenges encountered, the student may use

segments of one or the other language version of the test (e.g., by reading the test items or responding to the items in either English or the first language, or by using either language version to make sense of certain segments of text). Bilingual Glossary is also rated as highly sensitive to individual test takers' needs because it makes the translations of words available to the student when they are needed.

Test Taker Responses in Native Language and Directions Read in the Students' Native Language are rated as having a medium sensitivity to individual test takers' needs. Potentially, the individuals who provide these translation accommodations may be able, respectively, to properly interpret students' written responses and read the directions in the students' first language. However, in practice there is no way to ensure or to be certain that these individuals will have this ability.

Test Version in the Native Language and Directions Translated into Native Language are rated as having a low sensitivity to the individual test takers' needs because they are fixed formats that assume total proficiency in the first language and do not contain features that the student can use optionally when facing specific linguistic challenges.

Fidelity of Implementation. This dimension refers to how likely it is that a translation accommodation can be used as intended and in a standard form across all test takers. A common threat to fidelity of implementation of testing accommodations is that the individuals who administer a test to ELLs may interpret those accommodations in ways not intended.

In Table 4, Test Version in the Native Language, Directions Translated into Native Language, and Bilingual Glossary are rated as accommodations that can be implemented with high fidelity because their proper use is not shaped by the circumstances in which testing takes place.

Side-by-Side Bilingual Version of the Test is rated as an accommodation that can be implemented with a medium level of fidelity because, while the translation work involved is the same as Test Version in the Same Language, multiple formatting issues may compromise the quality with which this translation accommodation is provided. As an example, suppose that it is used with ELL students who are native speakers of Spanish. Once the assessment is available in English, the test is translated, and a double-sided page booklet is created in which the English and Spanish versions of each page appear respectively on the left and right sides. Unless the individuals in charge of developing the test in English and the individuals in charge of translating the test communicate effectively, the text space requirements in Spanish are unlikely to be properly addressed when the original version of the text in English is developed. As with many other languages, on average, words have more letters and sentences have more words in Spanish than in English (25% to 30% more). To make the longer Spanish text fit into the format imposed by the English version, the team in charge of assembling the side-by-side format may need to use smaller font sizes, narrower margins, and smaller spaces for the students' responses in the Spanish version. All these changes are unacceptable; they affect the equivalence of the language versions and are potentially highly detrimental to the ELL students' performance.

Test Taker Responses in Native Language and Directions Read in Native Language are rated as having a low fidelity of implementation because of the tremendous variability in the ways in which these accommodations can be provided to students.

Usability. This dimension refers to how likely it is that a translation accommodation can be used with ease by the test taker without making an effort other than using the knowledge and skills needed to respond correctly to a test. This dimension has to do with the extent to which the accommodation imposes extra cognitive demands or assumes in the test taker skills not related to

the assessment's content. Usability implies a cost in terms of the effort, learning, attention, or processing of information that the test taker needs to engage in, in order to be able to benefit from the accommodation.

In Table 4, Test Version in the Native Language and Directions Translated into Native Language are rated as having a high level of usability because they are delivered in a format that is familiar to all students—the same format as the English language version of the test.

Side-by-Side Bilingual version is rated as having a medium level of usability because, to benefit from this translation accommodation, students need to have certain meta-cognitive skills that allow them to both identify the portions of text they cannot understand in one language and look for those portions of text in the other language version. Also, while it is well known that navigating across languages is easier in side-by-side than top-bottom displays, the usability of the Side-by-Side accommodation may be limited in computer-administered testing, as the size of the screen may pose a limit to the amount of text that can be displayed in two languages simultaneously. Switching between screens in order to navigate across languages can place a tremendous cognitive demand on users—even those who are familiar with computers.

Bilingual Glossary is also rated as having a medium level of usability because it assumes in test takers the ability to recognize the words that they do not understand and the ability to perform alphabetical word searches. An exception to this limitation is the case of pop-up glossaries—an accommodation that is possible when tests are administered by computer and that allows the user to click on a word to see its translation in the first language displayed on the screen. A similar potential advantage is offered by audio glossaries—an accommodation that allows the student to click on a word to hear its translation. However, audio glossaries are a translation accommodation across both languages and language modes. While it is not harmful

for those students who do not need it, its use is justifiable only when there is certainty that students' listening proficiency in the target language is better than their reading proficiency in English.

An important issue related to the usability of Bilingual Glossary is that, in order to effectively serve its purpose, this accommodation has to be item-specific. That is, each word selected for inclusion in the glossary needs to be translated in the context of the item in which it appears because the same word in different items may have different meanings.

Test Taker Responses in Native Language and Directions Read in the Student's Native Language are rated as having a low level of usability. The former assumes that the ELL student writes proficiently in his first language; the latter assumes that the ELL student is more proficient in the listening mode in his first language than in the reading mode in English. Assuming total proficiency in the first language can be erroneous for the majority of ELL students with a history of schooling in English. In the case of directions read in the student's native language, the low usability of this translation accommodation also stems from the fact that many ELLs who are not in bilingual programs may feel uncomfortable using their first language in the school context and in a testing situation, and maybe with an individual with whom they may not be familiar.

Ensuring Validity and Fairness of Translation Testing Accommodations

As Table 4 shows, each translation accommodation has a different set of properties that makes it more or less likely to ensure valid and fair testing for ELL students. Deciding which type or types are to be used depends on factors such as the human resources available to implement or deliver the translation or the precision and accuracy of the information available about the characteristics of the students.

Translation accommodations with a low level of safety for untargeted test takers should not be used as blanket approaches with the entire population of ELL students. Even for students who are correctly classified as ELLs but have been schooled in English, testing them in their first language could be more harmful than beneficial. Translation accommodations with a low level of safety for untargeted test takers should be used only with specific groups of ELLs that are linguistically homogeneous and only with students who have a long history of schooling in their first language and no history or a short history of schooling in English. (Note 2).

Side-by-Side Bilingual Version of the test and Bilingual Glossary—rated as being highly sensitive to the individual test takers’ needs—are very promising translation accommodations. However, this attribute should not be taken for granted. A low quality of the translation may compromise the ability of the side-by-side bilingual version of the test to support students to make sense of the content of an item by navigating across language versions.

Overall, Bilingual Glossary, especially in the pop-up modality, appears to be the most promising translation accommodation, as it has the highest ratings on the four validity and fairness dimensions. However, to meet this potential, bilingual glossaries must be carefully constructed; they may fail to serve the ELLs’ needs with the level of specificity needed, if the words included in them have not been selected systematically and according to a solid conceptual framework on lexis (vocabulary), language proficiency, and academic language. Word frequency in English (a proxy for difficulty of text), criticality to learning or demonstrating knowledge of the topics assessed, and the identification of cognates (words that are morphologically and semantically similar in English and in the student’s first language) and false cognates (words that are morphologically similar but semantically different in English and in the

student's first language) are among the criteria that must be considered to determine which words should or should not be included in the glossary.

Of particular importance in the construction of glossaries is the need for ensuring proper contextualization. When used within a specific piece of text, a word “usually denotes meaning out of multiple meanings it inherently carries” (Dash, 2008, p. 21) because the user interprets the word according to the context in which it is used. In designing pop-up glossaries, contextualization can be easily accomplished because the translation of a target word appears next to it. The condition for effectiveness is that the translations of words be made “at the item level”—that is, the translation of a word should be determined according to the contextual space of the item. This translation needs to be made by humans. With the available current information technology, automatic translation of words cannot accomplish this level of sensitivity.

In printed glossaries, which are provided separately from the test materials, word translations with multiple meanings are decontextualized. If a word in English has multiple meanings, or if it can be translated in multiple ways in the ELLs native language, the glossary should provide appropriate clues for disambiguation of meaning.

The fidelity of implementation of translation accommodations that depend heavily on the characteristics of the individuals who provide them (test taker responses in native language and directions read in the student's native language) can be improved only if there is certainty that a sufficient number of qualified individuals will be selected and properly trained to participate in large-scale projects. However, implementation is likely to fail to the extent that it depends on the actions taken or decisions made by the individuals in charge of administering the tests. For successful test translation efforts in large-scale projects, it is better to assume uncertainty about the qualifications of the individuals who provide the accommodations.

Given the low rating of Test Taker Responses in Native Language and Directions Read in the Student's Native Language on the validity and fairness dimensions, these two translation accommodations should be considered as not valid and fair testing accommodations for ELLs. For this reason, the remainder of this document focuses on Test Version in the Native Language, Side-by-Side Bilingual Version of the Test, Directions Translated into Native Language, and Bilingual Glossary.

In the context of computer-based testing and the use of pop-up glossaries, usability may be an issue particularly important for ELL students, many of whom may not be familiar with computers and may require extra time to become familiar with clicking, mouse hovering, and other functions. The usability of pop-up glossaries as a form of test translation accommodation can be improved by providing the training needed to use it properly, allocating sufficient time for the students to become familiar with it, and improving its design.

Of special relevance to translation accommodations provided in computer-based testing is the need to keep the design of the interface as simple as possible.

An overwhelming body of evidence from the cognitive sciences and the realm of multi-media demonstrates that appealing stimuli that are not relevant to the task produce cognitive overload and, thus, make it difficult for the user to process the information provided (Clark & Feldon, 2005; Harp & Mayer, 1998; Mayer, Heiser, & Lonn, 2001). Hence, unnecessary visual elements, multiple navigation options, voices, avatars, and visual and sound effects constitute distracting factors that make it difficult for the test taker to make sense of items; unnecessarily increase cognitive demands; and threaten the validity of the test. Finally, because of the increased cognitive demands they impose, to be effective, Side-by-Side Bilingual Version of the Test and Bilingual Glossary need to be accompanied by an indirect linguistic support

accommodation—extended time allowed for test completion. A recent meta-analysis of testing accommodations concludes that promising testing accommodations such as pop-up glossaries are effective only when students are given generous amounts of time to complete their tests (Penock-Roman, & Rivera, 2011).

Assessment Translation Models

A translation model specifies the methods used to translate an assessment and to review and revise the translation of the assessment. These methods should be sensitive to both the characteristics of the items included in the assessment and the characteristics of the target population.

Important advances in the procedures for test translation have taken place in the last years, largely from experience with testing linguistically diverse populations in international test comparisons such as PISA and TIMSS (e.g., Grisay, 2003; Hambleton, 2005). Also, several guidelines for test translation have been generated for countries participating in those international test comparisons. In contrast, scant literature is available on translation procedures for linguistic minorities.

This dearth of literature should not be taken lightly. While there are important commonalities between translation in international test comparisons and translation in large-scale assessment in the U.S., it is important to keep in mind that the target populations in these two contexts are different. In addition, while necessary, translation guidelines are not sufficient in their effectiveness to address the complexities of test translation for ELL students. A rigorous procedure must be specified that addresses the heterogeneity of ELL populations that results, from, among other things, multiple patterns of bilingualism, multiple schooling histories, linguistic variation due to dialect and register within the same target language, and inaccurate or incomplete information about the language proficiency of ELLs in both their first and second language. The translation procedure should include a translation review component performed by independent translation evaluators.

Translation Team

The following professionals are essential to successful mathematics assessment translation for ELL students. Each of these professionals brings a unique kind of expertise to the translation team. Table 5 shows the required and desirable characteristics of these professionals.

- Mathematics teachers know the content area at the corresponding school level (e.g., elementary, middle school). They contribute expertise on the academic language in English and the linguistic challenges of constructing disciplinary knowledge.
- Translators have the theoretical foundation and formal training needed to address the technical aspects of translation. Translators should not be confused with or substituted for by linguists, English language arts teachers, specialists in literature in the target language, teachers of English as a foreign language, teachers of the target language as a foreign language, or writers. While the skills of those professionals are related to those of the translators, these skills are not the skills needed in test translation projects for ELL students. Translating from and into English entails a different set of skills. For example, a good Spanish-English translator does not make a good English-Spanish translator. Honoring this notion is regarded as sound, responsible practice in the translation profession. The American Translators Association (2012) issues different certificates for different combinations of languages.
- Bilingual teachers are familiar with the linguistic and cultural backgrounds of the ELL target population. They are familiar with the academic language used in mathematics textbooks in the target language and in the ELLs' communities and schools. Also, they are familiar with variation due to dialect in the target language and contribute

Table 5

Roles and Qualifications of Professionals Participating in the Assessment Translation: Mathematics

Professional	Form(s) of Participation	Required Qualifications			
		Credentials	Language Background	Cultural Background and Experience	Desirable Qualifications
Translator	<ul style="list-style-type: none"> • Pre-translation activities • Format design • Initial test translation • Translation reconciliation • Test translation review/revision • Follow-up activities 	Certified as an English-target language translator by a professional translators' organization or a higher education institution	Native speaker of the target language	Similar to the background of the ELL target population	Previous experience translating tests or documents in education, preferably in mathematics
Bilingual teachers	<ul style="list-style-type: none"> • Pre-translation activities • Word tagging • Format design • Cognitive interviews • Test translation review/revision • Follow-up activities 	Certified as a bilingual teacher by a higher education institution	Native speaker of the target language	Similar to the background of the ELL target population	Experience teaching mathematics
Mathematics teacher	<ul style="list-style-type: none"> • Word tagging • Format design • Cognitive interviews • Test translation review/revision 	Certified as a mathematics teacher in the corresponding school level (e.g., elementary, middle school) by a higher education institution		Experience teaching ELL students	User of the target language as either first or second language
Content specialist	<ul style="list-style-type: none"> • Word tagging • Format design • Test translation review/revision 	Mathematician or certified mathematics teacher with extensive curriculum development experience			User of the target language as either first or second language
Test developer	<ul style="list-style-type: none"> • Pre-translation activities • Word tagging • Format design • Translation review/revision • Follow-up activities 	Measurement specialist (psychometrician)			Experience in research and test development projects involving ELLs User of the target language as either first or second language
Sociolinguist	<ul style="list-style-type: none"> • Pre-translation activities • Word tagging • Format design • Test translation review/revision 	Ph.D. in sociolinguistics Specialty in bilingualism or experience with projects on bilingualism and bilingual populations			Experience in education projects User of the target language as either first or second language

knowledge of the vocabulary, syntactic structures, and discursive forms that are familiar to the majority of users of the target language. Bilingual teachers should not be confused with or replaced by teachers of English as a foreign language or teachers of the target language as a foreign language. Bilingual teachers are trained to address the functional aspects of communication among linguistic minorities. In contrast, teachers of English as a foreign language and teachers of the target language as a foreign language are trained to address the formal aspects of communication for students who are not linguistic minorities. Also, bilingual teachers should not be confused with or replaced by teachers who speak two languages. Bilingual teachers are teachers who, in addition to speaking two languages, have formal training in the teaching of bilingual populations.

- Content specialists inform the process of test translation on the linguistic aspects of the content of tests and the nature of the knowledge and skills assessed. Through discussion with colleagues, content specialists contribute to ensuring that the translation accommodation does not alter the intended meaning of items.
- Test developers facilitate the discussions of multidisciplinary teams in different stages of the translation process. They ensure that different perspectives of the translation team members are well taken into consideration and that the constructs assessed by the test items are preserved across languages.

Sociolinguists provide expertise on issues of language variation, language contact, bilingualism, and academic language. Sociolinguists should not be confused with or replaced by structural linguists, psycholinguists, sociologists, teachers of English as a foreign language, teachers of the target language as a foreign language, English language arts teachers, or specialists in English literature or target language literature. Only sociolinguists have the perspective needed to

understand communication from a social perspective and the complex interaction of language, dialect, and register as critical to encoding meaning in different languages.

Figures 4 to 7 show basic translation models for the four translation accommodations discussed in this framework. Some of the process components are common to all translation accommodations; other components are specific to one or two translation accommodations.

Translation Preparation Activities

As stated before, the process of test translation does not refer only to the activities directly related to translating text but also to those activities in assessment systems that affect the process of test translation. To ensure a sound process of test translation, project staff should develop adequate relationships with external agencies involved in the process of testing—mainly, contractors in charge of developing the tests in English and the officials who oversee the work of these contractors. The purpose of developing these relationships is to ensure that key professionals in the assessment system are aware of the ways in which the work of individuals not involved in test translation may actually affect test translation.

The fact that printed text in other languages takes more space than English is an excellent example that speaks to the importance of developing these relationships. Appropriate communication between the test developers of the original test in English and the translation project staff ensures that issues of formatting are worked out in a timely manner, at the very beginning of the process of test development.

An adequate timeline is perhaps the most important issue that needs to be worked out with those external agencies. Sound, valid test translation involves multiple professionals and several stages of development. Yet, unfortunately, test translation is still perceived by many as a task which one or two translators can complete in a couple of weeks. As part of developing

relationships with external agencies, the test translation project staff members need to support their colleagues from external agencies to understand the complexities of test translation and negotiate appropriate timelines for this work.

Because of the large numbers of test items that need to be developed, it is not reasonable to wait to initiate the process of test translation until all original items in English are available. Test translation project staff should be able to work out with external agencies a procedure by which they can gain access to the original items in English, so that the translation accommodations for them can be created as soon as they are generated in English.

Independent Translations and Translation Reconciliation

When appropriate, independent translators translate the test material, and a third translator—the translator reconciler—assembles a reconciled version of the translation, which is intended to resolve translation discrepancies and preserve meaning at the same level of difficulty across languages. To ensure a rigorous translation process, different translators should be involved in the different stages of the translation process.

Translation Review/Revision

In the four models, multidisciplinary teams make decisions about the characteristics of the translation accommodations and are responsible for creating the final versions of the translations. This is accomplished through translation review/revision sessions in which the team decides by consensus whether and how the translation should be improved.

For Test Version in the Native Language and Side-by-Side Bilingual Version of the Test, if the modifications are many or a substantial amount of the translation work has to be re-done, the translation may need one or several review/revision iterations (indicated by the dotted arrow). In

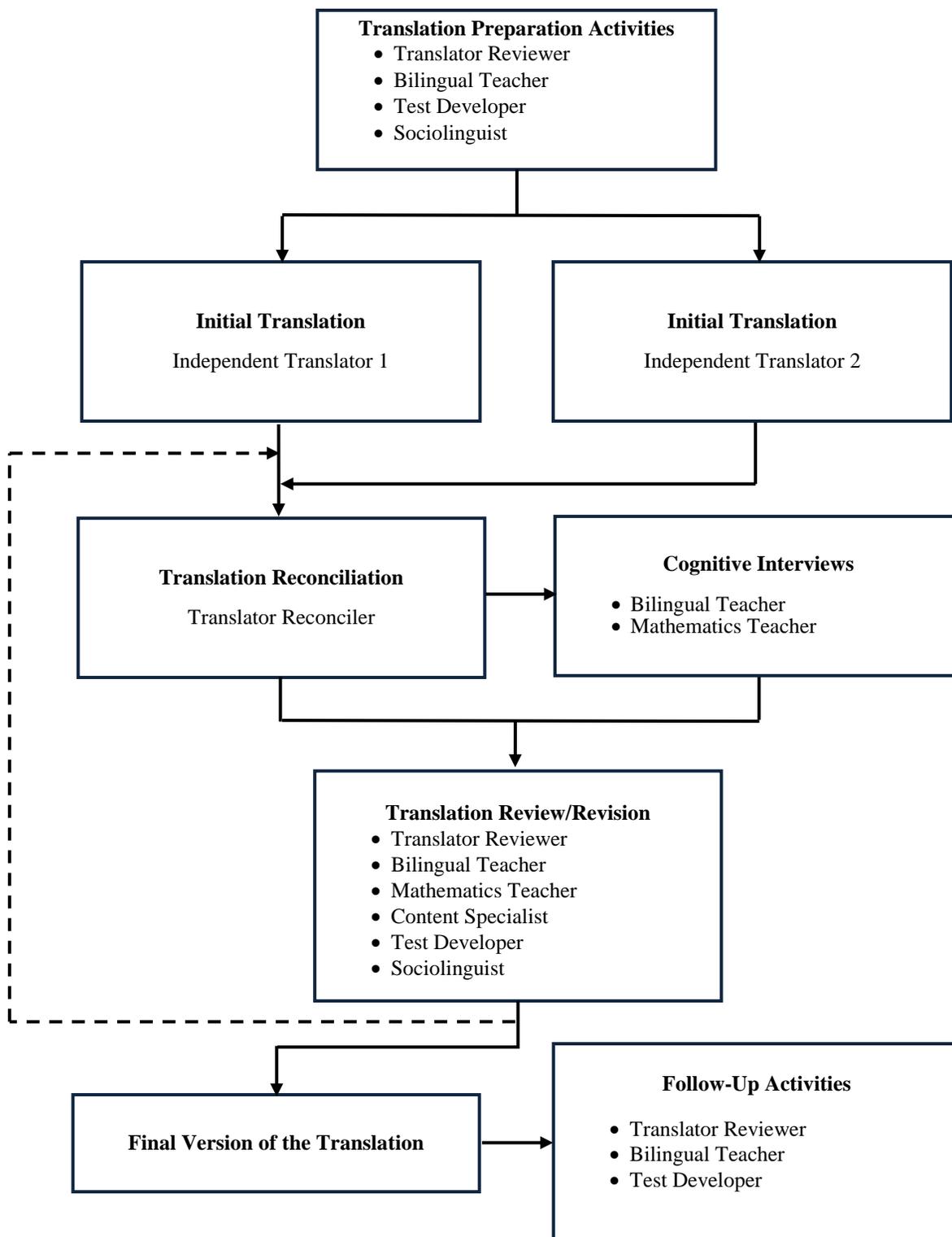


Figure 4. Translation model for the Test Version in the Native Language translation accommodation: Process components and professionals involved.

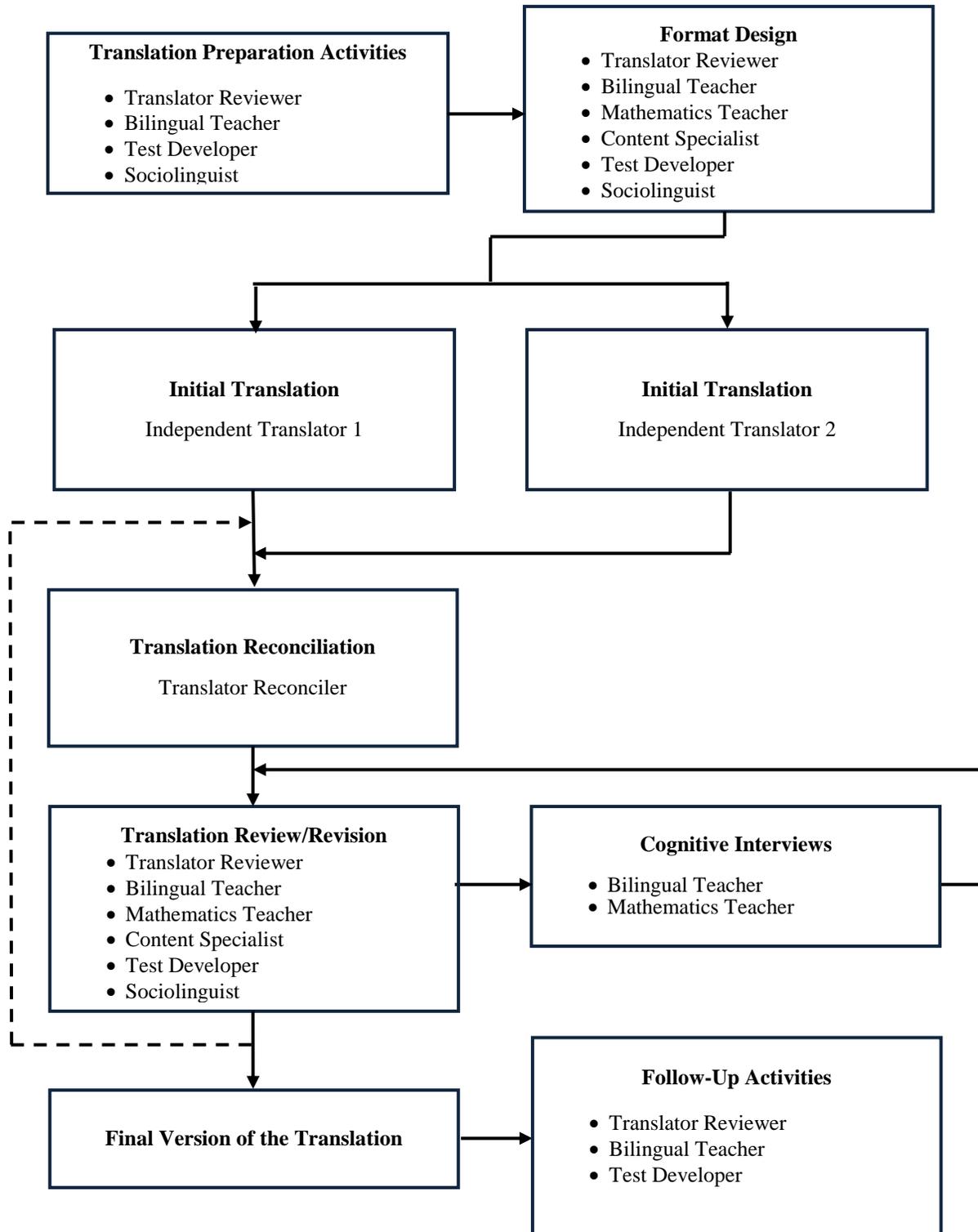


Figure 5. Translation model for the Side-by-Side Bilingual Version of the Test translation accommodations: Process components and professionals involved.

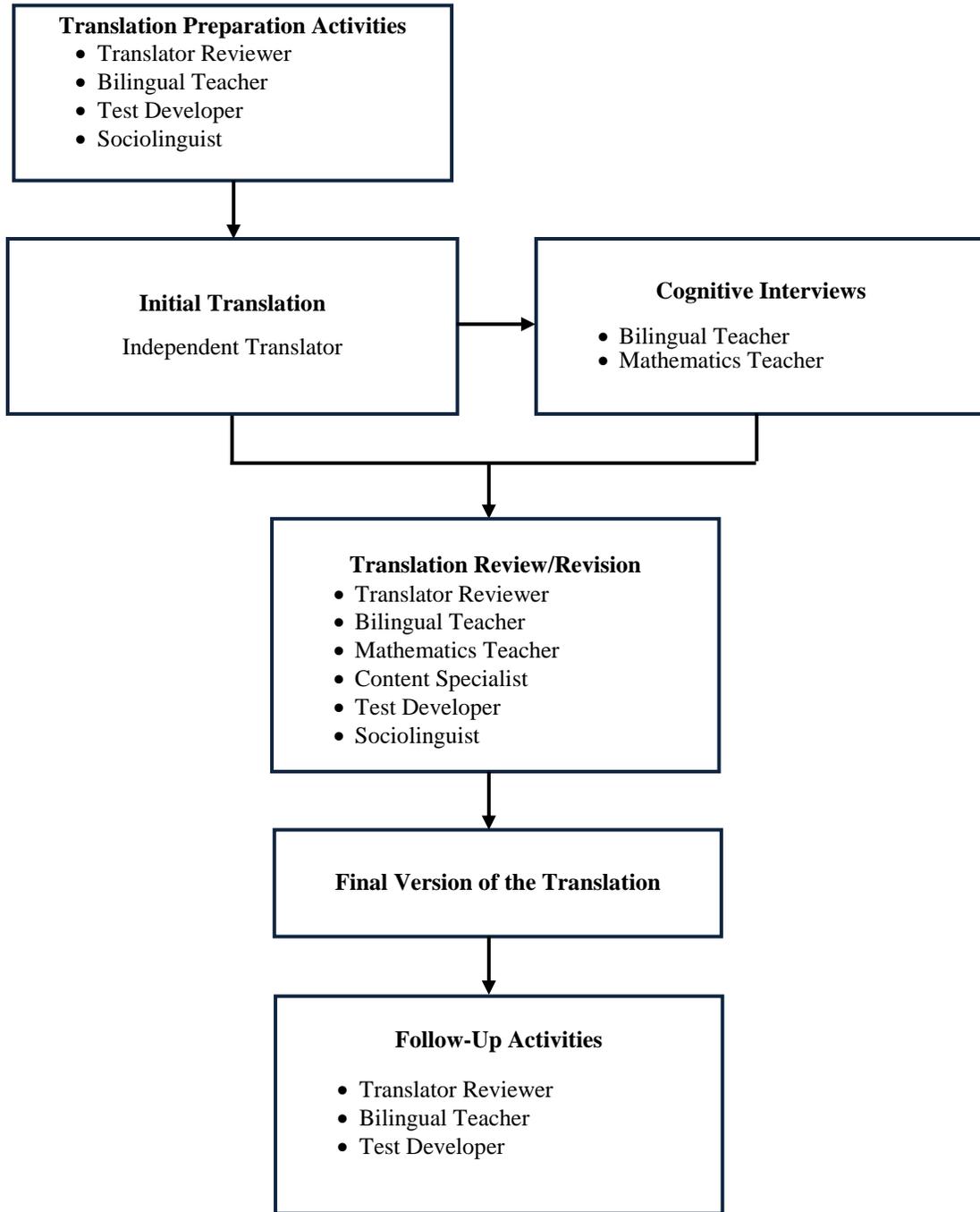


Figure 6. Translation model for the Directions Translated into Native Language translation accommodation: Process components and professionals involved.

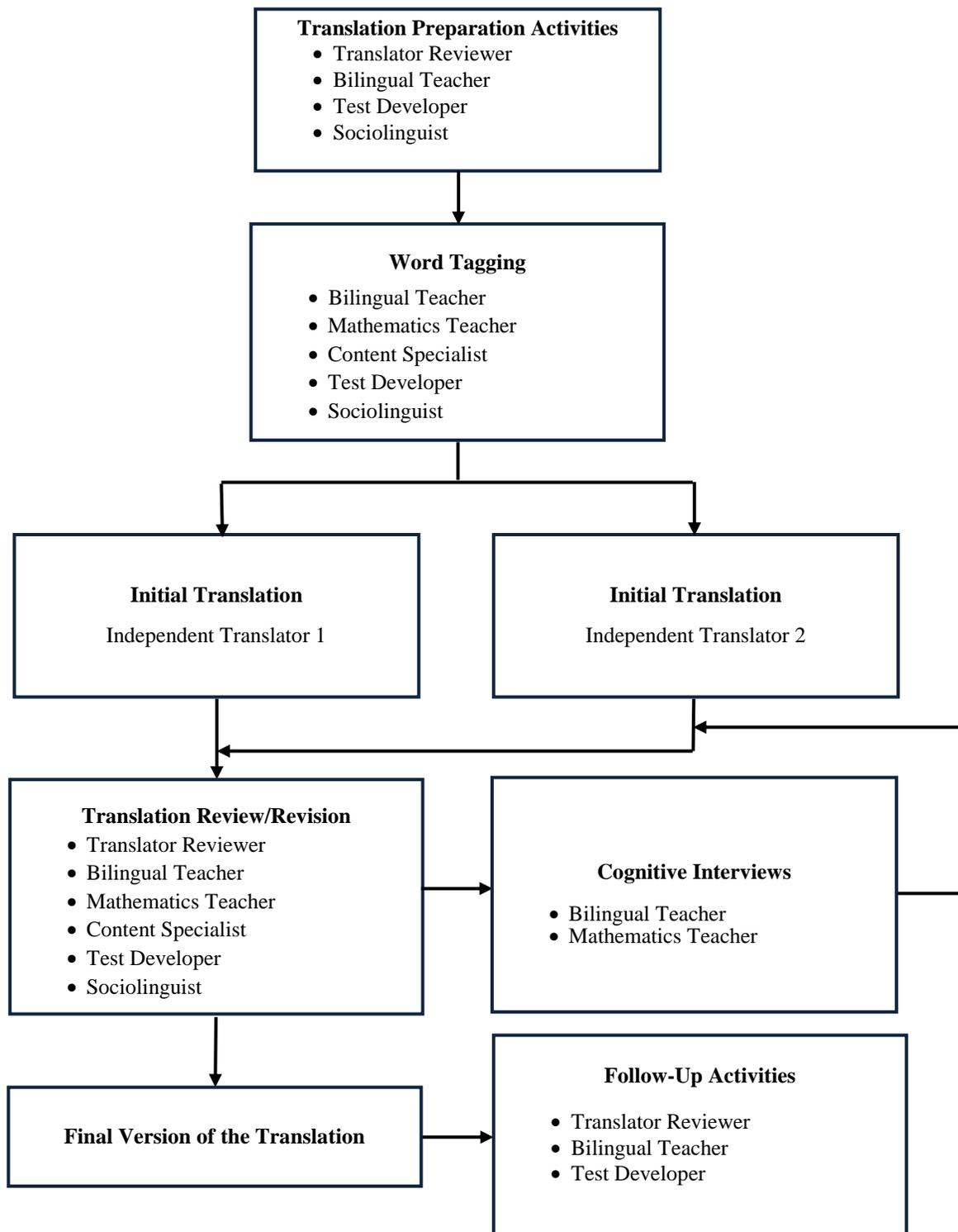


Figure 7. Translation model for the Bilingual Glossary translation accommodation: Process components and professionals involved.

these iterations, the translator reconciler refines the translation, and the translation review/revision team examines it and makes further refinements.

Figure 8 shows the translation review/revision procedures for each translation accommodation. As a first step in the review/revision procedures, and prior to comparing the test materials in the original version and in the accommodated version, the reviewers respond to the items as if they were students taking the test. This ensures that they become aware of the reasoning and knowledge involved in responding to the items and the ways in which the linguistic features of the items in the accommodated version influence their interpretations of the items. Decisions about possible modifications are reached by consensus, after each member has had the opportunity to propose and justify any modifications in the translation based on their own experience and professional background.

Note that the review/revision procedures are intended to be applied by item or by set of items with common directions. A discussion of the test materials “as a package” does not allow a discussion of the linguistic features of the test materials in detail.

Unlike translation review procedures used in international test comparisons, the review/revision procedures shown in Figure 8 focus on error (Solano-Flores, Backhoff, & Contreras-Niño, 2009). That is, translation review/revision members are instructed to find reasons that the translation may be flawed, rather than reasons that the translation is correct. There is evidence that this approach is more sensitive to subtle and sometimes important flaws in test translation.

Review/Revision procedure for the Test Version in the Native Language and Side-By-Side Bilingual Version test translation accommodations

- 1) The bilingual teacher, the translator, and other team members who can read in the target language:
 - independently read the translated item and respond to it as if each of them were a student taking the test;
 - independently compare the original and translated versions of the item and look for translation errors; and
 - independently edit the translated item (if needed) and write comments on it.
- 2) With facilitation from project staff, all team members discuss any proposed changes and decide by consensus whether and how the translation of the item should be modified.
- 3) Project staff keeps an updated copy of the translated item.

Review/Revision procedure for the Directions Translated Into Native Language test translation accommodation

- 1) The bilingual teacher, the translator, and other team members who can read in the target language:
 - independently read the (untranslated) items for which the directions apply and respond to them as if each of them were a student taking the test;
 - independently compare the original and translated versions of the directions and look for translation errors; and
 - independently edit the translated directions (if needed) and write comments on it.
- 2) With facilitation from project staff, all team members discuss any proposed changes and decide by consensus whether and how the translation of the directions should be modified.
- 3) Project staff keeps an updated copy of the translated directions.

Review/Revision procedure for the Bilingual Glossary test translation accommodation

- 1) The bilingual teacher, the translator, and other team members who can read in the target language:
 - independently examine the item in English and respond to it as if each of them were a student taking the test;
 - independently compare the target words in the original version and their translation in the glossaries and look for translation errors; and
 - independently change the translation of the target words (if needed).
- 2) With facilitation from project staff, all team members discuss any proposed changes and decide by consensus whether and how the translation of the target words should be modified.
- 3) Project staff keeps an updated copy of the translated target words.

Figure 8. Translation review/revision procedure for different test translation accommodations.

Cognitive Interviews

Cognitive interviews allow one to examine if the translation changes the construct the original language version intends to measure. In cognitive interview sessions, students read and respond to the translated items and are asked to verbalize or report their thinking, respectively as or after they respond to them. This “thinking aloud” allows the interviewer to identify whether the students interpret the translated items as intended (see Ericsson, 1993, for an original source on cognitive interviews). A follow-up short interview allows the interviewer to obtain additional information on the reasoning used by the students. There is a growing body of research on the use of cognitive interviews as an approach for validating tests (Baxter & Glaser, 1998; Hamilton, Nussbaum, & Snow, 1997; Ruiz-Primo, Shavelson, Li, & Schultz, 2001). Few, but important, documents exist that document the use of cognitive interviews as tools for examining how ELLs benefit from testing accommodations (e.g., Kachchaf, 2011; Kopriva, 2001).

Because cognitive interviews are costly and time consuming, it is not possible to conduct them for all translated items in large-scale assessment projects. Cognitive interviews need to be restricted to samples of items—those which contain the greatest amounts of text; include reading passages; or provide contextual information that is likely to pose translation challenges due to cultural differences, idiomatic expressions, and the like. In addition to contributing to improving the translation of the items selected, cognitive interviews can inform the translation process, as they may provide a clue to the adjustments needed in the translation process.

It is important to note that limited proficiency in English should not be regarded as an obstacle for ELL students to participate in cognitive interviews conducted *in English*. The majority of ELL students have the listening and speaking skills needed to interact in informal conversations in English. Indeed, there is evidence that, given the option to be interviewed in

English or in their native language, the majority of the ELL students prefer to be interviewed in English (Prosser & Solano-Flores, 2010). One reason is that ELLs' limited proficiency in English is more related to their development of academic language in English than their development of basic communication (conversational) skills in English. Another reason is that, for the majority of ELLs in the U.S., English is associated with the school context.

Cognitive interviews should also provide information relevant to determining the extra amount of time needed to allocate for ELL students to benefit from a translation accommodation. As mentioned before, to be effective, direct linguistic support accommodations may be accompanied by generous amount of extra time for test completion.

For Bilingual Glossary and Side-by-Side Bilingual Version of the Test, cognitive interviews are used not only to probe understanding of the translation but also to identify whether and how students benefit from (or struggle with) the design of these translation accommodations. Information collected on the way in which they use the accommodation to gain access to the content of items informs the review/revision of the translation.

Format Design

Test Version in the Native Language and Side-by-Side Bilingual Version of the Test translation entail the same type and amount of work—translating the text of the test in full. However, for Side-by-Side Bilingual Version of the Test, there is a stage for designing the format of the test. The importance of this stage should not be underestimated. Each target language and even the mathematics content of each grade has a specific set of features in the written and printed form that need to be addressed. The fact that Spanish text takes 25% to 30% more space than English (and which requires adequate coordination between test translators and the professionals in charge of developing the original items in English, so that the two language

versions have truly comparable formats) is a simple but powerful example of the design issues that need to be addressed. Another example related to test length is the translation of tables that contain text. How will the tables have to be adapted, so that the amount of text fits in the cells of the table? Yet another example is the letter “y,” which is used in mathematics to denote a variable (as in the y axis of a graph), and is also the conjunction “and” in Spanish. What graphic conventions will be used to make the difference evident to the student? The decisions on how to address these and many more design issues should not be left to editors, as these decisions need to be based on knowledge of the linguistic challenges of testing ELLs.

The design stage of the Side-by-Side Bilingual Version of the Test translation accommodation should be conducted with a sufficiently large sample of items (i.e., items of different types, grades, and topics). The participating professionals create a draft of the bilingual format and refine it as they encounter and resolve different formatting issues. This process should be done separately for every target language, as each language has a specific set of features in written and printed form.

Word Tagging

The first stage in the translation model for Bilingual Glossary consists of identifying the words that need to be translated. Through word tagging sessions, the word tagging team examines each item and identifies the terms that should be translated by translators to be included in the printed, pop-up, or audio glossary. An important part of this stage consists of developing a set of rules for deciding when a word should be included in the glossary. Should the glossary include only terms that are part of the contextual information provided by items? Should all the academic language terms be excluded? What criteria should be used to determine when a term counts as academic language? Should the terms with multiple meanings be included? What

are terms that are likely to pose a challenge to ELL students, or to users of the target language? Should both cognates and false cognates be included? How should English terms, which are used with different meanings in different items, be defined? These are issues that need to be discussed at length and should be resolved based on current knowledge on language and the disciplines. For example, there is evidence that many terms that are not exclusive of the register of a discipline (e.g., *therefore*) pose a challenge in the learning of that discipline simply because their frequency of use in everyday life contexts is low. Also, terms that are common in everyday life but whose meaning is slightly different in the context of a discipline may be challenging to students.

Automatic translation, even at the lexical level is likely to be flawed because it is not sensitive to context. The words tagged should be translated by human beings and should be specific to each item.

Final Version of the Translation

The final version of the translation is the translation given to publishers and other contractors for purposes of assembling the test. The final document should be accompanied by directions concerning formatting and other aspects of the production of the test.

Follow-Up Activities

The translation process does not end with the final version of the translated test materials. Other activities need to take place once the translated test materials are handed to other professionals involved in the development of tests. The project's staff needs to collaborate with publishers and other contractors in charge of assembling the test, printing it, or making it available for computer-based administration to ensure that the test, as it will be given to the students, has the intended characteristics.

The need for follow-up activities stems from the fact that the publishers' and other contractors' scopes of work do not necessarily address the characteristics of translated materials. As a consequence, their actions may not be entirely sensitive to the characteristics of the target language or the characteristics of ELLs. As an example, due to software incompatibilities, accents and other characters that are common in the target language but do not exist in English may be lost when the translation materials are transferred to the platform used for computer-based administration. Another example is that the team in charge of creating the computer user interface may decide to add graphic components, visual and sound effects, complex navigation features, voices, or avatars—features that increase the complexity of the items and threaten the validity of the translated accommodation.

These and many more issues difficult to anticipate may arise after the final version of the translation is delivered. Project staff needs to be able to be in continuous communication with publishers or other contractors and provide them, when necessary with feedback that ensures that the integrity of the translated test materials is preserved.

Systematic Development of Test Translation Accommodations

Especially for large translation projects in which massive numbers of items are translated and multiple translation teams participate, it is important to count on a set of conceptual tools and documents that ensure consistency in the process of test translation. Four elements are fundamental to systematically developing test translation accommodations: the use of assessment translation dimensions as a conceptual tool, the use of assessment translation specifications as a reference tool, the use of translation support materials, and an appropriate documentation of the translation process.

Assessment Translation Error Dimensions

As discussed earlier, in the context of large-scale assessment, and for the purposes of ELL testing, translation concerns not only translators' actions but also the actions of all individuals in an assessment system that can affect the process of test translation.

The inclusion of graphs in mathematics tests provides an example that illustrates the relevance of this expanded view of translation in ensuring fair and valid testing for ELLs. Figure 9 shows a multiple-choice item that asks the student to select the best description of the shape of a line in a graph. The correct option is B.

A slight distortion in the graph takes place while transferring or copying the electronic files or while formatting the text and the graphic material in the translated version. As a result of this slight distortion, in the translated version, the axis x ends up being just a little longer in proportion to the axis y than in the original version, as shown in Figure 10. Due to this distortion, students tested in the translated version may be more likely to select Option A as the correct option.

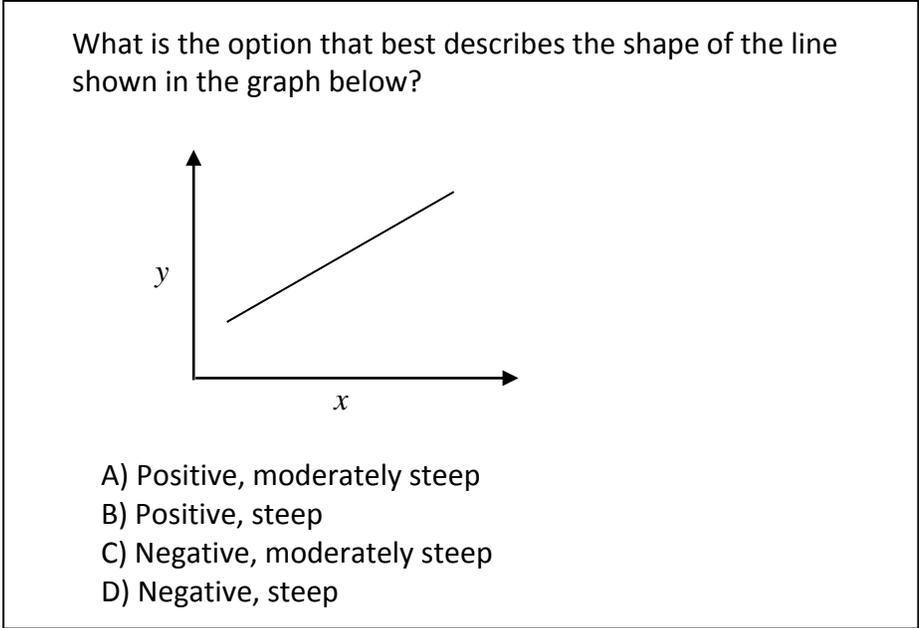


Figure 9. Original English version of an item.

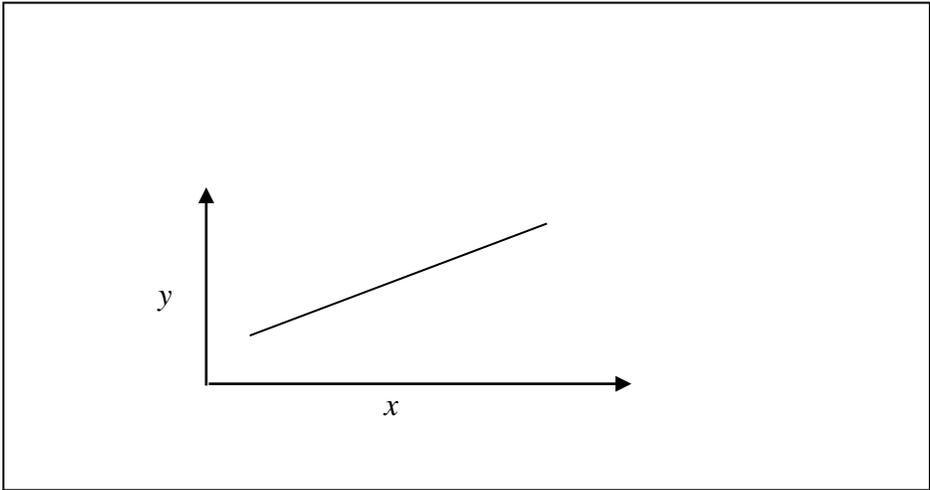


Figure 10. Graph shown in the translated version of the item.

In this example, the distortion of the image does not have to do with the translators' actions or with the linguistic features of the translation. Yet the item, in its translated version, has a serious error that compromises its validity.

For translation review/revision purposes, translation errors can be grouped according to translation error dimensions, as shown in Table 6. The set of types of errors included in each dimension may vary depending on the test translation context (e.g., international test comparisons or the testing of linguistic minorities) or the nature of the items examined (e.g., multiple-choice and constructed-response items). The examples listed are relevant to reviewing/revising test translation in the context of ELL assessment. While the list should not be regarded as exhaustive, it can be used as a document to help translation reviewers to examine different aspects of translation. However, it is important to note that the list shown is based on examining translations of tests into Spanish. Different errors may be identified when reviewing/revising translations into other languages.

An important notion for translation reviewers to keep in mind is that translation error is multidimensional (Solano-Flores, Backhoff, & Contreras-Niño, 2009). For example, a spelling mistake can also be an error that alters the original intended meaning of a sentence. Or, the literal (word by word) translation of a sentence can also alter the construct an item is intended to measure. Due to this multidimensionality, it is important that, in the translation review/revision sessions, the facilitator allow team members to discuss in depth the implications of the errors they detect.

*Table 6
Test Translation Error Dimensions with Examples. Adapted from Solano-Flores, Backhoff, & Contreras-Niño, 2009).*

Translation Error Dimension	Examples
Item Design Dimensions	
Style	<ul style="list-style-type: none"> • Unconventional use of accents, uppercase letters, and lowercase letters • Errors related to punctuation and spelling
Format	<ul style="list-style-type: none"> • Change of size, style, or position of graphs, tables, and illustrations • Change of font style and margins • Omission and insertion of graphic components
Conventions	<ul style="list-style-type: none"> • Grammatical inconsistency between options and between stem and options • Change in the order of options (in multiple choice items)
Language Dimensions	
Grammar and Syntax	<ul style="list-style-type: none"> • Literal translation • Wrong prepositions • Unconventional and unnatural syntactic structures • Collapsing sentences
Semantics	<ul style="list-style-type: none"> • Inappropriate adaptation and literal translation of idiomatic expressions • False cognates; alteration of meaning; insertion or omission of words • Use of imprecise terms or terms with multiple meanings
Register	<ul style="list-style-type: none"> • Literal translation of technical terms or translation of terms in ways that are unfamiliar to the ELL students in their first language, as it is used in the U.S.
Content Dimensions	
Information	<ul style="list-style-type: none"> • Inconsistent translation of the same term • Insertion or omission of terms and sentences • Change in the frequency with which key terms are used
Construct	<ul style="list-style-type: none"> • Omission, insertion, or inaccurate use of technical terms • Possible alteration of the item's cognitive demands or of the ways in which the content of the item is interpreted
Curriculum	<ul style="list-style-type: none"> • Discursive style of item not used in the curriculum
Origin	<ul style="list-style-type: none"> • More than one correct option • None of the options entirely correct • Bias: Misrepresentation of gender, racial or linguistic groups; situations that are unfamiliar to ELLs; etc.

Notice that two of the content dimensions, Curriculum and Origin refer to issues that can be detected but cannot be corrected during the translation review/revision sessions. Especially important is bias due to racial or gender stereotypes, the use of situations with which ELL students are unlikely to be unfamiliar, and the like (see Hambleton & Rodgers, 1995, for a list of potential sources of bias). It is not uncommon to detect errors in the original items when they are examined from the perspective of translation. Potential detection of such errors is another reason to allot adequate time to the translation/review process when initial timelines are established. Again, effective communication between the translation project staff and the developers of the assessment in its original version in English is important to ensure that these errors are addressed properly in both the original and translated versions of the items.

Assessment Translation Specifications

In large translation projects in which many items are translated, several sets of translators need to be hired and several teams need to be assembled respectively to perform the different activities for each translation accommodation. This multiplicity brings with it the challenge of ensuring standardization in the characteristics of the translation.

Assessment translation specifications are documents intended to ensure this standardization. They consist of sets of rules that establish the vocabulary and discursive style to be used across all translated items. Translation specifications ensure constancy in the characteristics of the translations regardless of the personal style and preferences of the individuals involved in the translation. Also, they optimize the efficiency of the work of both translators and members of the translation review/revision team.

Table 7 shows the appearance of a translation specifications document for translating mathematics items into Spanish. For the sake of simplicity, only two vocabulary entries and two

discursive style entries are shown. This document should be made available to all the professionals participating in the translation project. Needless to say, a translation specifications document needs to be developed for each target language.

Table 7
Examples of Entries in an Assessment Translation Specifications Document: Spanish

Vocabulary entries	Translation rule	Comments and Justification
<i>billion</i>	<i>mil millones</i>	Note than billón means, <i>a million millions</i> in Spanish.
<i>rectangle</i>	<i>rectángulo</i>	Do not use <i>rectángulo</i> to refer to the figure with four equal sides. In Spanish, <i>cuadrado</i> (<i>square</i>) is not a subset of the category, <i>rectángulo</i> (rectangle).
Discursive style entries		
Addressing the student	Use the familiar form, tú (e.g., <i>Mira la gráfica</i> , instead of <i>Mire la gráfica</i>)	Do not use the <i>usted</i> form, whose conjugation tends to be unfamiliar to Spanish speaking ELLs in the U.S.
Plural gender (e.g., the children, the students)	Use masculine plural (<i>los niños, los estudiantes</i>) to refer to plurals that include both males and females	Avoid forms such as <i>los niños y las niñas</i> or <i>los estudiantes y las estudiantes</i> , which are politically correct but increase the reading demands. Also avoid forms such as <i>el (los) estudiante(s)</i> , which are complex and difficult to read.

The translation specifications document should be developed before the translation process begins. However, it is important to recognize that specifications may keep evolving as experience translating items accrues. Because the entries and rules to be included depend on the characteristics of the items to be translated, project staff should translate a sample of items according to the procedures shown in Figures 4 to 7 (depending on the test translation accommodation intended). This sample should be representative of the different school grades and the different kinds of items included in the assessment (i.e., multiple-choice or open-ended;

with illustrations or without illustrations; different topics), as these different types of items for different grades are likely to have different linguistic characteristics. Project staff should keep a record of the issues encountered and the ways in which they are resolved. It is based on this experience that a translation specifications document like the one illustrated in Table 7 can be constructed.

Translation and Translation Support Materials

In order for the process of test translation to be effective, the professionals who participate in it must be provided with an appropriate set of translation support materials—documents needed to properly interpret the text translated or the text to translate. These translation support materials help participants to properly address the contextual aspects of language and make the necessary refinements concerning the type of knowledge assessed and the level of complexity of the language used in the corresponding school grade. These materials inform the discussions held by the teams of professionals who participate in the translation process.

Table 8 provides a list of translation and translation support materials that should be used. The original English version and the translation of the test (the translation materials) are obvious materials. English-target language-English dictionaries and content and specialty dictionaries in the target language can be used as reference materials respectively on the general and specialized, disciplinary use of terms. Instructional resources in English and instructional resources in the target language (e.g., internet instructional resources and textbooks) contribute to ensuring that the source and target language versions of the test are equivalent. The last four materials are information on the grade and topic assessed by each item, information on the standards and knowledge assessed by each item, assessment framework, and the common core

Table 8

Translation and Translation Support Materials to Be Made Available to Each Type of Participant in the Translation Process

	Word tagging team	Independent translator	Translator reconciler	Cognitive interviewer	Translation review/revision team
Original test in English	Yes	Yes	Yes	No	Yes
Translated test materials	---	---	Yes	Yes	Yes
English-target language-English dictionaries	No	Yes	Yes	No	Yes
Specialty dictionaries in the target language	No	Yes	Yes	No	Yes
Instructional resources in English	No	Yes	Yes	No	Yes
Instructional resources in the target language	No	Yes	Yes	No	Yes
Information on the grade and topic assessed by each item	Yes	Yes	Yes	Yes	Yes
Information on the standards and knowledge assessed by each item	Yes	No	Yes	Yes	Yes
Assessment framework	Yes	No	Yes	No	Yes
Common core standards	Yes	No	No	No	Yes

standards. These materials are useful for participants to ensure that the features of the translated tests do not alter the constructs measured or the intended level of difficulty of the items.

Different sets of materials should be made available to different professionals and teams of professionals, depending on the translation accommodation and the stage in the process of test translation. Needless to say, all the materials should be made available to the translation review/revision team, as this team decides on the characteristics of the final version of the translation.

Documenting the Process of Test Translation

The process of test translation should be documented with evidence of sound practice. More specifically, the following information should be provided:

- 1) Rationale justifying the translation accommodation selected to support fair and valid testing for ELL students, based on both knowledge (or uncertainty about) the characteristics of the ELL populations to be assessed and the extent to which the translation accommodation generated is likely to meet the four fairness and validity dimensions: safety of untargeted test takers, sensitivity to individual test takers' needs, fidelity of implementation, and usability.
- 2) Discussions of the actions taken to determine the amount of time for test completion appropriate to provide the translation accommodation selected.
- 3) Detailed information on the individuals who participate in the process of test development. More specifically, information on the extent to which their background, experience, and formal training meet the required qualifications (credentials, language background, and cultural background and experience) and desirable qualifications discussed in the framework.

- 4) Evidence that the stages of the process for developing the selected translation accommodation have been completed and the professionals have performed the corresponding activities indicated in the framework.
- 5) A strong rationale and appropriate evidence showing that sufficiently large and representative samples of items have been used for cognitive interviews.
- 6) A strong rationale and appropriate evidence showing that the samples of ELL students included in cognitive interviews are representative of the populations of ELLs with which the translation accommodation is to be used.
- 7) Copies of the translation specifications document and a list of the translation support materials made available to project participants.
- 8) Evidence that the test translation project staff has successfully performed translation preparation activities that ensure proper coordination with external agencies (mainly, contractors in charge of developing the original items in English and the officials who oversee the work of these contractors). As a result of this coordinated work: (a) the timelines for test translation should be commensurate with the magnitude of the work involved and the complexity of the test translation, (b) the translation project staff should have timely access to the original items developed in English, and (c) the process of development of the items in English should take into consideration format issues relevant to the administration of test translation accommodations in the target language.
- 9) Evidence that the test translation project staff has successfully performed follow up activities. More specifically, evidence that the test translation project staff has successfully collaborated with publishers and other contractors in charge of assembling the test,

printing it, or making it available for administration by computer, to ensure that the translation accommodation' integrity is preserved.

Final Remarks

Important challenges in the development of effective translation accommodations for English language learners are the tendency to underestimate the complexity of language and translation issues and the compartmentalization of activities in the process of test development. A systemic view of test translation allows practitioners and decision makers to appreciate the fact that the development of test translation accommodations does not start and end with the act of translating a test. Rather, the process involves the interaction of individuals in charge of translating test materials with colleagues who develop tests in the original language and with colleagues in charge of producing and publishing tests.

The individuals in charge of developing translation accommodations have both the responsibility to ensure that the process of test translation is sensitive to the complexity of language, translation, and linguistic groups and the responsibility to properly address the systemic components that influence test translation. The selection of qualified professionals, the establishment of adequate timelines, and coordinated work with external agencies involved in the process of testing are critical to successfully developing and using test translation accommodations for English language learners.

Notes

- Note 1. Glossaries should not be confused with customized dictionaries, in which the definitions are given in English and do not involve any form of translation.
- Note 2. A special group not considered in this framework is that of students who are native English speakers and attend bilingual, dual immersion programs in which some of the instruction they receive is in a language that is not English. While the relation between their first and second languages is the reverse of that for their ELL counterparts, these students are in no way equivalent to a linguistic minority. First, their first language is the same as the dominant language in the society in which they live. Second, their exposure to a second language through a bilingual program is optional. Thus, test translation accommodations for this group of students may not be relevant or effective.

References

- Abedi, J. (2004). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher*, 33 (1), 1-14.
- Abedi, J. (2006). Psychometric issues in the ELL assessment and special education eligibility. *Teacher's College Record*, 108(11), 2282-2303.
- Abedi, J. (2007a). English language learners with disabilities. In C. C. Laitusis & L. L. Cook (Ed.), *Large-scale assessment and accommodation: what works* (pp. 23-35). Arlington, VA: Council for Exceptional Children.
- Abedi, J. (2007b). English language proficiency assessment and accountability under the NCLB Title III: An overview. In J. Abedi (Ed.), *English language proficiency in the nation: Current status and future practice*. Davis, CA: University of California, Davis.
- Abedi, J., Lord, C., Hofstetter, C., & Baker, E. (2001). Impact of accommodation strategies on English language learners' test performance. *Educational Measurement: Issues and Practice*, 19(3), 16–26.
- American Educational Research Association (1999). *Standards for Educational and Psychological Testing*. Washington, DC: Author.
- American Translators Association (2012). ATA certification program: A guide to ATA certification. Retrieved, July 31, 2012, from http://www.atanet.org/certification/aboutcert_overview.php
- Artiles, A. J., & Ortiz, A. (Eds.). (2002). English language learners with special needs: Identification, placement, and instruction. Washington, DC: Center for Applied Linguistics.
- Baxter, G. P., & Glaser, R. (1998). Investigating the cognitive complexity of science assessments. *Educational Measurement: Issues and Practice*, 17(3), 37-45.

- Butler, Y. G., & Hakuta, J. (2006). Bilingualism and second language acquisition. In T. K. Bhatia & W. W. Richie (Eds.), *The handbook of bilingualism*. Malden, Ma, Blackwell Publishing, Ltd.
- Capps, R., Fix, M., Murray, J., Ost, J., Passel, J., & Herwanto, S. (2005). *The new demography of America's schools: Immigration and the No Child Left Behind Act*. Washington, DC: The Urban Institute.
- Clark, R. E., & Feldon, D. F. (2005). Five common but questionable principles of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning*. Cambridge: Cambridge University Press.
- Crystal, D. (1997). *The Cambridge encyclopedia of language* (2nd ed.), Cambridge, UK: Cambridge University Press.
- Dash, N. S. (2008). Context and contextual word meaning. *SKASE Journal of Theoretical Linguistics*, 5(2), 21-31.
- Durán, R. P. (2008). Assessing English-language learners' achievement. *Review of Educational Research*, 32, 292-327.
- Ercikan, K. (2002). Disentangling sources of differential item functioning in multi-language assessments. *International Journal of Testing*, 2, 199-215.
- Ericsson, K.A. & Simon, H.S. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, Massachusetts: The MIT Press.
- Francis, D., Rivera, M., Lesaux, N., Kieffer, M., & Rivera, H. (2006). *Practical guidelines for the education of English language learners: Research-based recommendations for the use of accommodations in large-scale assessments*. Portsmouth, NH: RMC Research Corporation,

Center on Instruction. Retrieved July 15, 2012, from <http://www2.ed.gov/about/inits/ed/lep-partnership/interventions.pdf>

- García, O., & Kleifgen, J. A. (2010). Educating emergent bilinguals: Policies, programs, and practices of English language learners. New York: Teachers College Press.
- Greenfield, P.M. (1997). You can't take it with you: Why ability assessments don't cross cultures. *American Psychologist*, *52* (10) 1115-1124.
- Grisay, A. (2003). Translation procedures in OECD/PISA 2000 international assessment. *Language Testing*, *20*(2), 225-240.
- Grosjean, F. (1985). The bilingual as a competent but specific speaker-hearer. *Journal of Multilingual and Multicultural Development*, *6*, 467-477.
- Halliday, M.A.K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. London, UK: Edward Arnold (Publishers), Ltd.
- Hambleton, R.K. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In R.K. Hambleton, P.F. Merenda, & C.D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Hambleton, R., & Rodgers, J. (1995). Item bias review. *Practical Assessment, Research & Evaluation*, *4*(6). Retrieved August 15, 2012 from <http://PAREonline.net/getvn.asp?v=4&n=6>
- Hamilton, L. S., Nussbaum, E. M., & Snow, R. E. (1997). Interview procedures for validating science assessments. *Applied Measurement in Education*, *10*, 181-200.
- Harkness, J. A., & Schoua-Glusberg, A. (1998). Questionnaires in translation. *ZUMA-Nachrichten Spezial*, January.

- Harry, B. & Klingner, J. (2006). *Why are so many minority students in special education?* New York: Teacher College Press.
- Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology*, 90, 414-434.
- Kachchaf, R. R. (2011). *Exploring problem solving strategies on multiple-choice items: Comparing native Spanish-speaking English language learners and mainstream monolinguals*. Doctoral dissertation, University of Colorado at Boulder.
- Kindler, A. L. (2002). Survey of the states' limited English proficient students and available educational programs and services 2000-2001 summary report. Office of English Language Acquisition, Language Enhancement and Academic Achievement for Limited English Proficient Students. Washington, D.C., National Clearinghouse for English Language Acquisition & Language Instruction Educational Programs.
- Klingner, J. K., & Solano-Flores, G. (2007). Cultural responsiveness in response-to-intervention models. In C. C. Laitusis & L. L. Cook (Eds.), *Large-scale assessment and accommodations: What works?* Arlington, VA: Council for Exceptional Children. (pp. 229-241).
- Kopriva, R. (2001). *ELL validity research designs for state academic assessments: An outline of five research designs evaluating the validity of large-scale assessments for English language learners and other test takers*. Paper prepared at the Council of Chief State School Officers Meeting, Houston, TX, June 22–23.
- Kopriva, R. J., Emick, J. E., Hipolito-Delgado, C. P., & Cameron, C. A. (2007). Do proper accommodation assignments make a difference? Examining the impact of improved decision making on scores for English language learners. *Educational Measurement: Issues and Practice*, 26 (3), 11-20.

- Mackey, W. F. (1962). The description of bilingualism. *Canadian Journal of Linguistics*, 7, 51-85.
- Mayer, R.E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology*, 93(1), 187-198.
- Migration Policy Institute (2010). *Top languages spoken by English language learners nationally and by state*. ELL Information Center Fact Sheet Series, No. 3. Retrieved July 31, 2012, from http://www.migrationinformation.org/ellinfo/FactSheet_ELL3.pdf
- NEA (2008). *English language learners face unique challenges*. NEA Policy Brief. Education Policy and Practice Department, NEA Human and Civil Rights Department, Center for Great Public Schools.
- Nettle, D., & Romaine, S. (2002). *Vanishing voices: The extinction of the world's languages*. New York, NY: Oxford University Press, Inc.
- No Child Left Behind Act of 2001 (NCLB), Pub. L. No. 107-110, § 115 Stat.1425 (2002).
- Oller, D. K., Pearson, B. Z., & Cobo-Lewis, A. B. (2007). Profile effects in early bilingual language and literacy. *Applied Psycholinguistics*, 28, 191-230.
- Penock-Roman, M., & Rivera, C. (2011). Mean effects of test accommodations for ELLs and non-ELLs: A meta-analysis of experimental studies. *Educational Measurement: Issues and Practice*, 30(3), 10-28.
- Prosser, R. R., & Solano-Flores, G. (2010). *Including English language learners in the process of test development: a study on instrument linguistic adaptation for cognitive validity*. Paper presented at the Annual Conference of the National Council of Measurement in Education, Denver, Colorado, April 29-May 3.

- Rivera, C., Collum, E., Schafer, L., & Sia, J. K., Jr. (2006). An analysis of state assessment policies regarding the accommodation of English Language Learners. In C. Rivera, & Collum, E. (Ed.), *State assessment policy and practice for English language learners* (pp. 1–173). Mahwah, NJ: Lawrence Erlbaum Associates.
- Ruiz-Primo, M. A., Shavelson, R. J., Li, M., & Schultz, S. E. (2001). On the validity of cognitive interpretations of scores from alternative concept-mapping techniques. *Educational Assessment*, 7(2), 99-141.
- Sireci, S. G., Scarpati, S. E., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research*, 75(4), 457-490.
- Solano-Flores, G. (2008). Who is given tests in what language by whom, when, and where? The need for probabilistic views of language in the testing of English language learners. *Educational Researcher*, 37(4), 189-199.
- Solano-Flores, G. (2009). The testing of English language learners as a stochastic process: Population misspecification, measurement error, and overgeneralization. In K. Ercikan & W. M. Roth (Eds.), *Generalizing from Educational Research* (pp.33-48). New York: Routledge.
- Solano-Flores, G. (2011). Language issues in mathematics and the assessment of English language learners. In K. Tellez, J. Moschkovich, & M. Civil (Eds.), *Latinos/as and mathematics education: Research on Learning and Teaching in Classrooms and Communities*, (pp. 283-314). Charlotte, NC: Information Age Publishing.
- Solano-Flores, G., Backhoff, E., & Contreras-Niño, L.A. (2009). Theory of test translation error. *International Journal of Testing*, 9, 78-91.

- Solano-Flores, G., & Gustafson, M. (2013). Assessment of English language learners: A critical, probabilistic, systemic view. In M. Simon, K. Ercikan, & M. Rousseau (Eds.), *Improving Large Scale Assessment in Education: Theory, Issues, and Practice*. Taylor & Francis: Routledge.
- Solano-Flores, G., & Li, M. (2006). The use of generalizability (G) theory in the testing of linguistic minorities. *Educational Measurement: Issues and Practice* 25(1), 13-22.
- Solano-Flores, G., Li, M., Speroni, C., Rodriguez, J., Basterra, M., & Dovholuk, G. (2007). *Comparing the properties of teacher-adapted and linguistically-simplified test items for English language learners*. Paper presented at the annual meeting of the American Educational Research Association. Chicago, IL. April 9-13.
- Stansfield, C. W. (2003). Test translation and adaptation in public education in the USA. *Language Testing*, 20, 188–206.
- Thurlow, M. L. (2007). State policies and accommodations: Issues and implications. In C. C. Laitusis & L. L. Cook (Ed.), *Large-scale assessment and accommodation: what works* (pp. 13-22). Arlington, VA: Council for Exceptional Children.
- Thurlow, M. L., Elliott, J. E., & Ysseldyke, J. E. (2003). *Testing students with disabilities: Practical strategies for complying with district and state requirements* (2nd ed.). Thousand Oaks, CA: Corwin.
- Valdés, G., & Figueroa, R. A. (1994). *Bilingualism and testing: A special case of bias*. Norwood, NJ: Ablex.

van de Vijver, F.J.R. & Poortinga, Y.H. (2005). Conceptual and methodological issues in adapting tests. In R.K. Hambleton, P.F. Merenda, & C.D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Wolfram, W., Adger, C. T., & Christian, D. (1999). *Dialects in schools and communities*. Mahwah, NJ: Lawrence Erlbaum Associates.

Young, J., & King, T. C. (2008). *Testing accommodations for English language learners: A review of state and district policies*. College Board Research Report No. 2008-6. New York: The College Board.