



10 Things You Should Not Forget When Adapting a Test Into Spanish

A Practical Decalogue

Ricardo Rosas^{ORCID}, Marcelo Pizarro^{ORCID}, and Victoria Espinoza^{ORCID}

Centro de Desarrollo de Tecnologías de Inclusión (CEDETI-UC), Escuela de Psicología, Pontificia Universidad Católica de Chile, Santiago, Chile

Abstract: Adapting a test from one language to another is no easy task. The literal translation of the items is certainly not enough since the meanings and references change from one language to another. In the case of Spanish adaptations, the problem becomes even more complex, as there are many varieties of Spanish, almost as many as Spanish-speaking countries. This article presents 10 tips for adapting tests to Spanish regarding the morphological, syntactic, and pragmatic elements that should be considered when adapting an instrument to Spanish. These tips are based on experience with the adaptation of various instruments, such as the Wechsler Intelligence Scales (WISC-III, WAIS-IV, and WISC-V) and the Big Five from English, and the ELFE scale, MARKO-D, and the Berlin Intelligence Test from German.

Keywords: assessment, adaptation, translation

Frege (1892), in his famous text *Über Sinn und Bedeutung* (translated as *Sense and Reference* in English and as *Sentido y Referencia* in Spanish), generated the controversy which is the subject of his entire work. The essential question of the text is, what is the difference between sense and meaning? He posed this question based on the example of how we refer to Venus as the morning star or the evening star. Undoubtedly, both expressions allude to the same star, that is, to the same reference. But the sense of the morning star and evening star is quite different. What is interesting in this case is that in other languages, such as English or Spanish, there was intense controversy over the translation of the words Sinn and Bedeutung or the meaning that the German philosopher wanted to give to both. So much so that the translators of both languages finally had to submit to the connotation that the German philosopher gave to the word Bedeutung. Thus, in neither of the two languages did the famous work take on the usual meaning of dictionary – *significado* in Spanish and *meaning* in English – having to change it to *referencia*, or *denotación* in Spanish, and *reference* in English.

This brief philosophical disquisition represents the task of cross-cultural adaptation of psychometric tests. The job of test adapters is to rescue the original sense, even if the reference changes. The above example illustrates the difficulties of translating the original meaning from one

language to another when words in different languages have different semantic fields.

Often, as we will see throughout this article, the main task of the test adaptor is to decide whether the original test sense or the reference is retained. Usually, for reasons of economy, it is a matter of maintaining the reference. But more often than not, we must try to rescue and preserve the sense. A great example of this is the decision we made regarding the adaptation of an item from the WISC-III Figure Completion scale in the early 2000s. The original item was a piano with missing black keys. This item was a low-difficulty item in the original test, as the figure presented is a piano found in all public schools in the United States. However, in Latin America, the piano is a relatively rare instrument, restricted to the elites, and is rarely present in public schools. Including the piano would certainly have generated an extreme bias by socioeconomic level, which is undoubtedly a problem in the measurement of intelligence. Not being able then to include the item reference, the question arises as to what is the objective of the item. And this answers the question, is there any local musical instrument frequently used and present in Latin American schools in general and Chilean schools in particular? The answer to this question is the guitar. And to make sense of the original question even more

precisely, a new question arises: Does the guitar have any very striking and salient element that, if omitted, would be easily recognizable by children? The answer is the tuning pegs. The results of this adaptation allowed the item to be used with a Chilean sample.

As seen in this example, preserving meaning, a guiding principle for adapting foreign tests, is equivalent, in psychometric terms, to protecting content validity. Suppose the content validity of the original test is assured. In that case, the other psychometric properties, such as evidence of construct validity, the validity of relationships with other variables, and reliability, are estimated on solid conceptual and empirical grounds.

As a research team, our experience in adapting tests to Spanish comes mainly from German and North American tests. Among the former, the Berliner Intelligenzstrukturmodell (Jäger, 1982; Kleine & Jäger, 1987; Rosas, 1990; Süß & Beauducel, 2005), the Mathematische Rechenkonzepte (Marko-D, Fritz et al., 2017; Rosas & Espinoza, in preparation), and the Leseverständnistest für Elementarschüler (ELFE-II, Lenhard et al., 2017; Rosas, Escobar, and Espinoza, in preparation) stand out. Among the latter, we have the Wechsler scales for children and adults WISC-III (Ramírez & Rosas, 2007), WISC-V (Rosas, Pizarro, et al., 2022; Wechsler, 2014), WAIS-IV (Rosas et al., 2014; Wechsler, 2008), and the Hearts and Flowers Test for measuring executive functions in children (Davidson et al., 2006; Diamond, 2013; Rosas, Espinoza, et al., 2022). Finally, we created a personality test based on the Big Five model from the public IPIP database (Goldberg, 1999; Goldberg et al., 2006; Rosas & Pizarro, in preparation).

The adaptation of each of these instruments has involved challenges in detecting the best way to preserve the original sense of each of the items and their instructions. The linguistic and cultural differences in Spanish-speaking countries affect all the dimensions of language when adapting a test from another language: phonological, morphosyntactic, semantic, and, finally, pragmatic aspects.

Our approach to this Decalogue will be essentially practical and based on real examples of situations that have occurred to us when adapting to Chilean Spanish, and sometimes when trying to maintain equivalent tests with other countries in the region, such as Colombia, Ecuador, and Uruguay.

Each Spanish-Speaking Country Is Different, Even in Speech

Some foreigners tend to think that Latin America and Spain because they share language and culture, also share, in

Frege's words, sense, and references. However, this assumption is not entirely true. As we shall see in this article, we can distinguish, at least for the purposes of psychometric test adaptation and following Henríquez Ureña (1921) division, five different regions in America: (1) bilingual regions from North America, Mexico, and republics such as Guatemala, El Salvador, Nicaragua, Costa Rica, and Panama; (2) Antillas españolas (Cuba, Puerto Rico, and República Dominicana); and (3) Andean Regions (Venezuela, Colombia, Ecuador, Peru, and Bolivia), Río de la Plata (Argentina, Uruguay, and Paraguay), and Chile.

This grouping is due to historical reasons, or linguistic reasons, or both (Ángel & Pacheco, 2014; Cockcroft et al., 2015; Levy & Schady, 2013; Rocha Carpiuc, 2016). For example, as we will see in the next point of the Decalogue, the countries of the Río de la Plata, Uruguay, and Argentina always deserve special and differentiated attention, as they are the only Latin American countries that use *voseo* in a standard way, even in written language. This has important consequences for the design of test instructions, as children in these countries will find instructions in standard Spanish strange. The same is true throughout Latin America with standardized tests in Spain, since in Spain the normal use of the second person plural is the colloquial form (*vosotros*), whereas throughout Latin America it is the formal form (*ustedes*). But even grouping countries in this way is not enough to think that tests can easily be standardized across national borders. And this is because in Latin America there is a great cultural diversity between countries, which is usually exacerbated and emphasized in schools, as it is often anchored in national episodes that gave rise to independence from Spain or other heroic war deeds that become very important factors in the national identity of each Latin American country. And these episodes or national characteristics are those that tend to be important when determining the relevant content to be included, for example, in the information subtest of the Wechsler scales, or in any scale that investigates the average knowledge of what a given child is expected to have learned at school.

It should be remembered that Spanish-speaking people mainly inhabit two continents, Europe and America. And these two territories have physical, climatic, and cultural characteristics that generate very important differences in lifestyles, objects of daily use, flora, and fauna in the habitat and a long list of etcetera. For this reason, it is not uncommon for even commonly used words to have very different meanings in different American countries and in Spain. It should be noted that this is not a phenomenon exclusive to Spanish, as there is evidence that this is the case for English, which is distributed as a native or second primary language in at least four continents. One study even shows that the factor structure of intelligence measured in WAIS-III

Table 1. Differences of three common words between Uruguay and Chile

English word	Spain	Uruguay	Chile
Pineapple	piña	ananá	piña
Swing	columpio	hamaca	columpio
Fridge	nevera	heladera	refrigerador

differs when comparing English and South African students (Cockcroft et al., 2015).

For example, when adapting the Chilean version of the WISC-V to Ecuadorian Spanish, we noticed that in the Similarities subtest, one of the items included was banana, which in Chile, as in all European countries and North America, has a univocal meaning and refers to the sweet, yellow, delicious fruit whose peel is the subject of endless illustrations of people slipping on it. This word was adapted from the original North American test. However, in developing the Ecuadorian version of the test, we discovered that in Ecuador, the word “plátano” (banana) refers to the green banana, which is a fruit normally used for cooking and eaten salted. What in the rest of the world is known as banana, in Ecuador is known as *guineo*.

Several similar examples occurred with the Chilean adaptation of the original Uruguayan translation of the ELFE test mentioned above: Although Uruguay and Chile are highly similar countries in education and culture, there are frequently used words that differ in meaning and usage (see examples in Table 1).

Two Spanish-Speaking Countries Are Particularly Different

In addition to phonetic or phonological differences, there are other differences in the orality of the language that may affect the assessment conditions. We refer specifically to the prosodic variations between different native Spanish-speaking countries. In different countries, and even in different cities within a country, differences in intonation, stress, and rhythm of speech can be observed. These differences do not usually represent difficulties for the general understanding of the language. However, there are some elements that may interfere with the specific understanding of assessment instructions or may make it necessary to incorporate differences in the translation or creation of items in a given assessment instrument. We refer specifically to what is known as Rioplatense Spanish.

Rioplatense Spanish refers to the form of speech of the Spanish language used in some areas of Argentina (Buenos Aires, the southern provinces of Santa Fe, Entre Ríos, and Patagonia) and in a large part of Uruguay (di Tullio & Kailuweit, 2011). *Voseo* is one of the morphosyntactic

phenomena that characterizes Rio de la Plata Spanish because *voseo* affects verbs. *Voseo* is used to express the relationships of lesser distance and informality (Palacios, 2016). *Voseo* considers the replacement of the pronoun “tú” by the pronoun “vos” associated with a verbal conjugation characterized by a change in the traditional accentuation, and in some cases by morphological changes in the words, for example, “vos mirás” (you look), “vos comés” (you eat), and “vos jugás” (you play). In English, *thou* is the most equivalent word to *voseo*, although it is considered an Archaism that has been replaced with the word *you* in almost all contexts. Only certain parts of England and Scotland use it, although rather scarcely.

In the case of test instructions for children, forms of interaction that involve less social distance from the test taker are often used, so “usted” is replaced by “tú.” In the case of Rioplatense Spanish, *voseo* is used, which constitutes a challenge when adapting or developing instructions and test items that can be used in a general way in Latin America.

Specifically in the case of ELFE II, the initial translation was developed in Uruguay (von Hagen, 2018) using *voseo* as an integral part of both items and written instructions. During adaptation for standardization in Chile, the use of “vos” was changed to “tú” and the verbs were converted to their traditional form without *voseo*. This adaption allows its proper application in Chile and other parts of Latin America. The final adaptation of ELFE II (Rosas, Escobar, and Espinoza, in preparation) considers the unification of speech registers to make the test as transversal as possible. This is done with the possibility of having general standardized versions of the tests that can be applied in more than one country in the region in mind.

In this scenario, it is necessary to analyze whether these changes with respect to *voseo* could affect the performance of students who speak Rioplatense and Costa Rican Spanish, since it could generate a distance between the children and the evaluators. It is important to reflect on this aspect by establishing an analysis of the students’ results to generate a fair assessment space for all students, regardless of the region they come from.

To avoid this tradeoff between specific and general versions, modifying only the instructions and keeping the same items is possible. This makes international standardization possible, even sharing norms, but with different application protocols for voicing and nonvoicing countries.

Even Isolated Graphemes Can Generate Differences in the Spanish-Speaking World

There is a variety in the way certain graphemes are named in Spanish. For example, in Spain, the grapheme “b” is

called *be* and the grapheme “v” is called *uve*. Although both are pronounced in the same way in speech, in writing, a distinction is made between the two. In Latin America, these graphemes have different names: The “b” is referred to in different places as *be*, *be larga*, and *be alta* and the “v,” *ve*, *ve corta*. As in Spain, these distinctions do not affect speech either, as both graphemes correspond to the same phoneme.

Differences in grapheme’s name can impact certain tasks in some cognitive assessments, such as the Letter and Number Retention subtests of the Wechsler scales. In these subtests, a series of letters and numbers are presented, which the test takers must sort according to two rules: The numbers must be ordered from lowest to highest and the letters must be ordered alphabetically. In this case, the inclusion of either *be* or *ve* grapheme could be problematic since they are pronounced identically, but they are located in quite differently in the alphabet. The following example illustrates this issue:

L – B – R – 1 – A – 6.

In this case, a Latin American examiner would read this item as follows: /ele/, /be/, /erre/, /uno/, /a/, and /seis/. The pronunciation of the second stimulus (B) is the same as that of the V. Since they are pronounced in the same way, a test taker would not know for sure whether what they heard corresponds to B or V, so they could answer as follows (Table 2).

In both cases, the test taker’s answers are correctly ordered according to what they heard. However, the second assumption, in which the test taker believes that they heard V, is inconsistent with the stimulus, so they should receive 0 points. To avoid this type of confusion in this subtest, B and V should be avoided when items are presented in alphabetical order.

In line with the above, other graphemes have different names depending on the country, such as *y*. In some countries, this grapheme is called *y griega* and in others *ye*. It is important to consider that the first form is the most widespread and that the second began to be used in the mid-19th century to follow the rule of adding an *e* after a consonant. This generates a somewhat paradoxical situation, since the most well-known form of naming the grapheme, *y griega*, has the consequence that this stimulus has a much greater number of syllables than the others and consequently generates an imbalance in the difficulty of each stimulus. On the other hand, the simplest form of

pronouncing this grapheme, *ye*, is much less widespread, so it may not be recognized by all speakers.

The above considerations do not preclude the existence of a unique form of subtests, such as Letter and Number Retention, as most Spanish graphemes have a very similar pronunciation. It is important to be aware of some of these differences to avoid confusion. That is why in the adaptations we have made of the Wechsler scales we have changed all the *be* and *ve* to *de*, as well as changing all the Greek *y*’s to *equis* (x).

The Name of the Original Test Is Not Necessarily Retained

The name of tests that are gold standards cannot be translated or adapted. This is the case of the Wechsler scales, which retain their original names not only in Spanish-speaking countries, despite the difficulty and complexity of pronouncing them in Spanish: WISC, WAIS, and WPPSI. It should also be noted that for most Latin and Spanish psychologists, English is the second most widely used language, so the acronym in English makes perfect sense and is understood by all.

The question arises whether we can use the same principle for acronyms in other languages. We have had to make this decision in the case of three tests which we have translated from German: the BIS (Berliner Intelligenzstrukturmodell test), the MARKO-D (Mathematische Rechenkonzepte), and the ELFE (Ein Leseverständnistest für Elementarschüler). In the case of the first one, we decided to change the original acronym to its translation: MEIB (Berlin Intelligence Structure Model; Rosas, 1990), since the word BIS has its own meaning in Spanish that lends itself to misunderstandings as a test name. The other two have been left as the original, as the acronyms sound good in Spanish and do not have a literal meaning. In other words, if the name or acronym in the original language means something else in the language of adaptation, we suggest changing the name. This is a heuristic based on custom and common sense and should be evaluated on a case-by-case basis.

A second issue related to the name of the test in Spanish is its prosody, that is, how it sounds in our language. For example, when adapting the test Hearts and flowers by Adele Diamond (Davidson et al., 2006; Wright & Diamond, 2014) one of the six scales of the Yellow Red test (Rosas et al., 2022), we noticed that the Spanish translation (*Corazones y Flores or Flores y Corazones*) has almost twice as many syllables and does not flow naturally. For this reason, we decided to change it to cat/dog (*Gato-Perro*), which also meant changing the visual stimuli, but the test is essentially the same.

Table 2. Assumptions for letter-number sequencing

Examinee assumptions	Examinee answers
Examinee assumes B	A – B – L – R – 1 – 6
Examinee assumes V	A – L – R – V – 1 – 6

Finally, the Binding subtest (another of the six scales of the Yellow Red test), we decided to change it to Nexos, since in Spanish the expression Binding is read with a different pronunciation than in English, which does not make sense, while the expression nexos means exactly what the test measures.

Finally, in relation to the names, it is necessary to pay attention to the empirical results of the Spanish adaptation. When adapting the Big Five test from the item pool of the International Personality Item Pool, it turned out that the items most relevant in Spanish to two of the five main scales were best grouped under labels slightly different from those of the original. Thus, the original Responsibility scale was changed to Tidiness and the Openness to Experience factor was changed to Innovation.

Be Especially Careful With Language in Mathematics Instruction

Language impacts understanding in several areas of learning, and mathematics is one of them. Mathematics employs formal, abstract language, which demands precision for the accurate expression of specific concepts (Powell et al., 2019). The inadequate use of mathematical language can generate difficulties in the understanding of mathematical concepts, which is why it is necessary to know and use precise concepts with respect to the specific knowledge that is expected to be developed (Puga Peña et al., 2016).

This is also fundamental in the field of assessing mathematical competences, as an error at the linguistic level can modify the content to be assessed and therefore the performance expected of students. Thus, both when translating a test from a foreign language and when writing the questions directly, it is essential to review and use the specific and appropriate concepts for the assessment of each specific content.

In the case of the translation and adaptation of the MARKO-D test (Fritz et al., 2017) for the assessment of initial mathematical competences, a translation from German into Spanish was carried out, followed by a comparative review between the Spanish and German versions. Both the translation and the comparative review were carried out by native speakers of each target

language. The person who carried out the comparative review was also knowledgeable about the theoretical model underpinning the test.

In the comparative review process, the reviewer noticed a linguistic difference that modified the level of development assessed by some items. Specifically, there were two items that had the same verbal structure. The difficulty was found in the use in German of the word “größer,” which in Spanish translates indistinctly as “mayor” (greater) or “más grande” (bigger).

The initial item proposed in German was geared to the assessment of quantification, i.e., it considers the understanding of the relationship between a number and the quantity it represents. The expression “two plus x” requires an understanding of the number as a quantity with the possibility of making the necessary additive combinations. However, when translated into Spanish, the use of language determined a change in the content assessed, since by using the expression “greater than,” the question focused on the assessment of the ordinal function of numbers. This content corresponds to a level prior to quantification within the developmental progression of early mathematics. The expression “two numbers greater than 4” implies an understanding of the ordinal function of numbers, but not necessarily an understanding of the relationship between number and quantity. In contrast, the expression “two numbers bigger than 4” requires an understanding of the quantifier function of numbers. The translations and re-translations made can be seen in Table 3.

Vocabulary Tests Are Particularly Sensitive to Frege

Generally, it is the verbal tests that require the greatest degree of attention for their adaptation. The vocabulary subtest of the Wechsler batteries is a good example of this. In this subtest, a word is read to the examinee who must provide a definition that shows a good understanding of the word.

Adapting the vocabulary subtest to Spanish requires certain precautions because the direct translation of each item from one language to another could have an impact on the difficulty. Words are used in different contexts and situations in each language. For example, the word *glass*

Table 3. Translation stage for mathematical items in Marko-D

Translation stage	Item	Content assessed
Original German item	Wie heisst die Zahl, die um 2 größer ist als 4?	Cardinality
First Spanish translation	¿Cómo se llama el número que es 2 mayor que 4?	Ordinality
Back translation to German	Wie heißt die Zahl, die 2 größer als 4 ist?	Ordinalidad
Final Spanish translation	¿Cómo se llama el número que es dos más grande que 4?	Cardinality

can be translated as *vaso* or as *copa* depending on the context. It is possible for an English speaker to say *I drank a glass of water in the morning* or *I want a glass of wine in the afternoon*. The same word (*glass*) is used correctly for both contexts, although they refer to different types of liquid recipients. The direct translation for glass in Spanish changes depending on the context: The liquid recipients for drinking water and wine are called *vaso* and *copa*, respectively.

The effect that the context can have in the meaning of words has made necessary the use of cross-linguistic corpora for translation purposes (see Hansen-Schirra et al., 2013). This is particularly important in the translation from English to Spanish – and especially in cognitive assessment context – because words that are considered difficult, such as the ones listed in the Academic Word List, are used very frequently in Spanish because this language has its roots in Latin (Bushong, 2004). Therefore, it is necessary to consider the register of use of each word to obtain information regarding not only how frequently a word is used, but also the context in which it is used (Biber, 1993).

Rosas, Tenorio, and Pizarro (Rosas et al., 2014; Rosas, Pizarro, et al., 2022) used the Corpus of Contemporary American English (Brigham Young University, 2012) to determine the frequencies and register of use of each WAIS-IV and WISC-V word. In general, it was observed that simpler items had higher frequencies of word use than the more difficult items. Additionally, variation in the register of use was observed, with the first words being used most frequently in oral registers, the next words in popular written registers (e.g., newspapers, general information magazines), then in more specialized written registers (e.g., novels, literary texts), and finally, the words toward the end of the subtest were used in highly specialized written contexts (e.g., academic texts). This variation in registers of use was consistent with an increase in difficulty as there was a progression from simpler registers, such as orality, toward greater specialization in academic texts.

Once we had analyzed the frequencies and registers of use of each word in English, we translated each of the words into Spanish. Then, we carried out the same analysis of frequency and register of use of the translated words. For each case, we analyzed whether the words had a similar frequency and were used in the same registers as the original words, for which we used the “Corpus de palabras del español chileno” (Sadowsky

& Martínez-Gamboa, 2012). Although there are other frequency corpus for Spanish, such as the one available from the Real Academia de la Lengua Española (<https://corpus.rae.es/lfrecuencias.html>), these often do not include information regarding the register of use, which is necessary for comparing word equivalence in cross-linguistic analysis (Hansen-Schirra et al., 2013). In these cases, there were important differences in frequency or register of use, and we changed the word to one that had the same grammatical class as the original word, as well as similar frequencies and registers of use. In case the translated word had a similar frequency and usage register as the original word, then the word was kept for the Spanish version.

An example of the process is shown below. It is important to note that the words included do not correspond to items from the Wechsler scales. The purpose of the following example is purely demonstrative. Suppose a test in English has the following three words: *rocket*, *specie*, and *virtual*. By analyzing each of these words using the “Corpus of Contemporary American English” and comparing them with those of a Spanish corpus, the following results are obtained (Table 4).

The three examples illustrate different situations for adaptation. In the case of *rocket*, it can be seen that both English and Spanish use this word more frequently in the context of news. However, the difference in frequency is very different in one language than in the other, as in English this word appears around 20 times more often than in Spanish. This difference is evidence that this word is more common for an English speaker than for a Spanish speaker. In this case, we would replace the word *rocket* with a word that has a frequency more similar to the English word.

In the case of *specie*, the highest usage register is the same in both languages and the frequency of use is quite similar. In this case, the directly translated word could be used, since there are no major differences between the two languages.

Finally, the word *virtual* is an interesting case, since not only are there important differences in frequency (in Spanish, this word is used 45% more than in English), but the register of highest use is also different: While in English the word *virtual* is used more frequently in academic registers, in Spanish, it appears more frequently in magazines. The difference in registers suggests that this word in Spanish is more easily accessible to a Spanish speaker

Table 4. Differences in register

English word	English frequency	Most used register in English	Spanish frequency	Most used register in Spanish
Rocket	26,167	News	1,252	News
Specie	72,542	Academic	62,157	Academic
Virtual	19,729	Academic	34,390	Magazines

than to an English speaker, since magazines are more widely read than academic texts. Given the difference in both frequency and register of use, we would replace this word with another word that has a similar frequency and register of use to the original English word.

Another interesting case occurred when adapting the BIS vocabulary depth test. The original German test was the classic presentation of a word followed by four response alternatives, one of which represented the meaning of the word. The test was carefully designed and highly tested on the Berlin secondary school population. As a power test, the items were rigorously laid out in order of increasing difficulty. At that time (1980s), there was not yet the development of linguistic corpora as there is today, which allow the recommendations we have just made for vocabulary tests of the Wechsler scales to be followed. Therefore, the words were translated and tested in the same order with Chilean university students. The result? A total disaster, as all the words considered difficult in German were of Latin origin, the so-called *Fremdwörter*. These words are mostly commonly used in Spanish-speaking countries, which forced the authors to completely redo the test for Chile.

One Adapted Saying Is Better Than a 100 Mistranslated Ones

The Wechsler scales include sayings in the Comprehension subtests. In these items, the respondent is presented with a saying and asked to explain its meaning. The adaptation of such items requires special care, as the mere translation of a saying from one language to another may result in a sentence that is not recognized as a saying or is simply not comprehensible in any way.

Let us take the following example: A bird in the hand is worth two in the bush. The literal translation of this saying into Spanish is “un pájaro en la mano vale dos en el arbusto.” This literal translation does not make sense to a Spanish speaker since it will not be recognized as a proverb and because the final part of the translation (“vale dos en el arbusto”) does not make sense word for word. Although the translation could be corrected to make it more understandable to a Spanish speaker (e.g., *un pájaro en la mano vale el doble que uno en el arbusto*), it would not be recognized as a proper saying either.

For adapting this type of item, we have opted to look for equivalent Spanish proverbs. In the case of our example item, we would have chosen to ask for the Spanish equivalent of the same saying, which is “Más vale (un) pájaro en mano que cien volando” (a bird in the hand is better than a 100 in the air). This proverb refers to the same content as the original and is indeed a Spanish-language proverb.

Even Nonverbal Aspects of Tests Often Require Adaptation

Generally, nonverbal subtests do not require further item adaptation. Tests such as the Raven’s Progressive Matrices have evidence of validity across languages and cultures. However, there are items in certain subtests that may present some difficulties in adaptation, such as the Figure Completion (WISC-III) or Incomplete Figures (WAIS-IV) subtests. In these subtests, a picture is presented in which something is missing that the test taker must identify. Most of the items in these pictures present objects or situations that are common in Western culture. However, there are some items whose images present objects that are not as common in one culture as in another. The example of the piano and the guitar was already presented in the introduction.

Something similar occurred in the WAIS-IV adaptation. One of the items in this subtest showed an electric cooker. This type of cooker is commonly used in the United States, so it is a familiar object to people in that country. In the pilot phase of WAIS-IV in Chile, it was observed that none of the test takers recognized this object. In Chile, most cookers are gas cookers and only in recent years have electric cookers become more common, although only as a cooktop rather than as a stand-alone appliance. Given the differences between the two types of cookers, it was necessary to make an adaptation to this item, which consisted of replacing the electric cooker with a gas cooker. This change allowed the participants to recognize the object presented and, consequently, to give an answer.

These examples show when a cultural adaptation is mandatory, as they represent objects or customs that are explicitly asked about. There are other examples where this is not the case, but still merit adaptation: We refer to cultural objects that support the narrative of an instrument, although they do not directly influence the construct being measured. For example, the original MARKO-D instrument bases its entire narrative on two squirrels in the forest. Most of the items are built on questions of cardinality, addition and subtraction of acorns and flowers. Although knowledge of squirrels does not affect the mathematical operation, as these animals are nonexistent and unknown in Chile, they had to be replaced by dogs (the South African version of this instrument replaced them by meerkats, the Brazilian version of capuchin monkeys).

Cultural Uses Require Special Attention

Culture strongly influences the way people understand how their environment works. This has consequences for what is identified as intelligent behavior. Some subtests, such as the Comprehension of the Wechsler scales, deal precisely

with knowledge about functioning in one's own culture. In this subtest, examiners ask questions such as why should we wear clothes or what are watches for? These questions, as the examples illustrate, deal with topics that in most cases can be translated without major difficulties, since in all Spanish-speaking countries people wear clothes and know how watches work. However, there are some topics that arouse greater interest in English-speaking countries, such as the United States and the United Kingdom. Examples of this are topics related to the functioning of justice systems or those related to space exploration.

Regarding justice systems, the functioning of the legal system has clear differences between English-speaking countries and other countries. In particular, the jury system, in which a group of people is chosen to make a final decision in certain trials, allows citizens to have a greater knowledge and connection with the functioning of this system. In contrast, in Latin American countries, trial decisions are made by judges; in no case are decisions made by a jury made up of citizens.

The greater citizen participation in the functioning of the justice system in English-speaking countries allows us to understand the reasons for including questions on legal issues in the Comprehension subtests of the Wechsler scales. In contrast, in Latin American countries, justice systems do not consider the participation of citizens as jurors. This, perhaps, allows us to understand the reasons why questions referring to legal issues do not have good psychometric performances in the Spanish adaptations. An example of this occurred in the adaptation of the WAIS-IV scale in Chile because in the piloting stage we directly translated all the items, including those with legal themes. When scoring these items using the original correction criteria, we observed that none of the participants managed to obtain even 1 point on these items. As a consequence, the items referring to this topic were not included in the final version of the Chilean version.

Other topics in which cultural differences impact the performance of items in the Comprehension subtest are those related to spatial exploration. Space exploration has an important place in North American society, as they have been one of the pioneer countries in this type of research. For this reason, questions related to space exploration are included in the Comprehension subtests. This type of question is not suitable for Latin American countries, as it is not a topic that is familiar or close to the inhabitants.

Other subtests of the Wechsler scales in which it is important to pay special attention to culture are the Information subtests. This subtest includes general knowledge questions that have specific answers. One question that could be included in this subtest is, Who is Shakespeare (note: this question is not included in any of the Wechsler scales but serves for purposes of illustrating the

type of adaptation we have made). This question has a certain difficulty for an English speaker, which is seen in the relative position of this item in the subtest. This level of difficulty is determined by the frequency of correct and incorrect answers (a higher frequency of correct answers is indicative of an easier item; a lower frequency of correct answers is indicative of a more difficult item).

The difficulty of our example item (Who is Shakespeare?) can vary greatly between English and Spanish speakers. Shakespeare is typically recognized as the greatest, or one of the greatest, writers of the English language. That is why he has studied in English-speaking schools. In contrast, this author does not have the same level of recognition in Spanish-speaking countries. In Latin America, authors who write in Spanish are studied whether they are Latin American or from Spain. In this sense, Miguel de Cervantes is the Spanish author who has the greatest equivalence with Shakespeare, since he has studied in both Latin American and Spanish schools.

In this sense, a good adaptation of our example item (Who is Shakespeare?) would be Who is Miguel de Cervantes? In this example, the original topic, that is, writers widely recognized within a culture, is retained, modifying the question to adapt it to the target culture.

A final example of the adaptations we have had to make as a result of the impact of culture is with respect to one of the WISC-III information items. The original version of this battery included an item asking why it is important to wear a seat belt when traveling by car. This question makes sense when a culture has a practice of wearing seat belts, which is also given by automobile use. The United States has had a strong automobile culture from the 1950s onward, so the everyday experience of traveling by car is common, especially when a significant percentage of the population has access to a car. In contrast, when the WISC-III was adapted in Chile, the automobile was not as common as it is today, and there was no regulation on the use of seat belts. These factors meant that the psychometric qualities of this item were not adequate, which made it necessary to replace it.

Some Things Do Not Happen in Exactly the Same Way in the Northern and Southern Hemispheres

Many natural phenomena are affected by the hemisphere we inhabit. For example, in the northern hemisphere, water flows in eddies from left to right, and in the southern hemisphere, it flows from right to left. This phenomenon, known as Coriolis, causes the same thing to happen with the direction of rotation of hurricanes and waterspouts. We see marked differences in the sky: The constellation Ursa

Major is only visible in the northern hemisphere and the opposite is true for the Southern Cross.

The seasons of the year are also different in both hemispheres. When it is summer in the south, it is winter in the north. This means that Christmas in the southern hemisphere usually occurs at around 30 degrees Celsius. The strength of tradition means that Santa Claus comes to the children thickly wrapped around a Christmas tree festooned with artificial snow.

For these reasons, great care must be taken to adapt items that are in some way anchored in knowledge determined by the hemisphere one inhabits. One of the most notable examples we have observed occurred when standardizing the WISC-III picture arrangement test for Chile, which, as has been said, was based on the Spanish version adapted in Argentina. The most difficult item of the test consists of ordering, according to the shadow cast by a house in the Arctic, the passing of the hours of the day.

The sun rises in the east and sets in the west in both hemispheres, but the shadows move clockwise in the north and counterclockwise in the south. During the day in the northern hemisphere, the sun will reach its maximum position facing south because it will follow the direction of the equator (south). In the southern hemisphere, the opposite is true, as the sun reaches its maximum point in a northerly direction. And this makes the correct ordering of history in the northern hemisphere exactly the opposite of the southern hemisphere.

The amazing thing is that the Argentinian version of the WISC-III, from where this item was taken, always kept as correct only the arrangement that came from the Spanish test, i.e., the correct answer in the Northern Hemisphere!

Conclusions

The 10 tips presented in this article illustrate certain issues, topics, and problems we have encountered in adapting English instruments over the last 30 years. The way in which we have solved the difficulties has always been in pursuit of a fairer assessment situation for Spanish-speaking test takers. This is to follow globally recognized assessment guidelines, such as the standards (American Educational Research Association et al., 1999) and the ITC guidelines (Bartram et al., 2018).

We hope that these 10 tips will be helpful for other research and development teams to develop or adapt tests to Spanish. This will not only allow us to have a greater number of tests for the Hispanic population, but also in the generation of better instruments that are more accurate, provide more useful information, and that allow us to correctly assess those to whom we test takers ultimately owe it to ourselves: our own test takers.

Fortunately, Venus is also the morning and evening star in the southern hemisphere. In this, our beloved Gottlob Frege retains sense and reference.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *The standards for educational and psychological testing*.
- Ángel, M., & Pacheco, Q. (2014). División dialectal del español de América según sus hablantes: Análisis dialectológico perceptual [Spanish dialectal division according to their speakers: Perceptual dialectological analysis]. *Boletín de Filología XLIX*, 2, 257–309.
- Bartram, D., Berberoglu, G., Grégoire, J., Hambleton, R., Muniz, J., & van de Vijver, F. (2018). ITC guidelines for translating and adapting tests (2nd ed.). *International Journal of Testing*, 18(2), 101–134. <https://doi.org/10.1080/15305058.2017.1398166>
- Biber, D. (1993). Using register-diversified corpora for general language studies. *Computational Linguistics*, 19(2), 219–241.
- Brigham Young University. (2019). *Corpus of Contemporary American English*. Retrieved December 15, 2022 from <https://www.english-corpora.org/coca/>
- Bushong, R. W. (2004). *The academic word list reorganized for Spanish-speaking English language learners*. STARS Citation. <http://library.ucf.edu>
- Cockcroft, K., Alloway, T., Copello, E., & Milligan, R. (2015). A cross-cultural comparison between South African and British students on the Wechsler Adult Intelligence Scales Third Edition (WAIS-III). *Frontiers in Psychology*, 6, Article 297. <https://doi.org/10.3389/fpsyg.2015.00297>
- Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, 44(11), 2037–2078. <https://doi.org/10.1016/j.neuropsychologia.2006.02.006>
- di Tullio, A., & Kailuweit, R. (2011). *El español rioplatense: Lengua, literatura, expresiones culturales* [Rioplatense Spanish: language, literature, cultural expressions]. Iberoamericana/Vervuert.
- Diamond, A. (2013). Executive functions. In *Annual review of psychology* (Vol. 64, pp. 135–168). Annual Reviews. <https://doi.org/10.1146/annurev-psych-113011-143750>
- Frege, G. (1892). Über Sinn und Bedeutung [Sense and Reference]. *Zeitschrift Für Philosophie Und Philosophische Kritik*, 100(1), 25–50.
- Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, de Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe* (Vol. 7, pp. 7–28). Tilburg University Press.
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40(1), 84–96. <https://doi.org/10.1016/j.jrp.2005.08.007>
- Hansen-Schirra, S., Neumann, S., & Steiner, E. (2013). Cross-linguistic corpora for the study of translations. In *Insights from the language pair English-German*. De Gruyter Mouton. <https://doi.org/10.1515/9783110260328>
- Henríquez Ureña, P. (1921). Observaciones sobre el español de América [Observations on the Spanish in America]. *Revista de Filología Española*, VIII, 357–390.

- Jäger, A. O. (1982). Intelligenzstrukturforschung: Konkurrierende Modelle, Neue Entwicklungen, Perspektiven [Intelligence structure research: Competing models, new developments, perspectives]. In A. O. Jäger (ed.), *Berliner Beiträge Zur Intelligenzforschung* [Berlin contributions to intelligence research] (pp. 8–34). Freie Universität Berlin Forschungsprojektschwerpunkt Produktives Denken, Intelligentes Verhalten.
- Kleine, D., & Jäger, A. O. (1987). Replikation des Berliner Intelligenzstrukturmodells (BIS) bei brasilianischen Schülern und Studenten [Replication of the Berlin Intelligence Structure (BIS) model among Brazilian high school and university students]. *Diagnostica*, 33(1), 14–29.
- Levy, S., & Schady, N. (2013). Latin America's social policy challenge: Education, social insurance, redistribution. *Journal of Economic Perspectives*, 27(2), 193–218. <https://doi.org/10.1257/jep.27.2.193>.
- Palacios, A. (2016). *Dialectos del Español de América: Chile, Río de la Plata y Paraguay* [Dialects of American Spanish: Chile, Río de la Plata and Paraguay]. Routledge.
- Powell, S. R., Stevens, E. A., & Hughes, E. M. (2019). Math language in middle school: Be more specific. *TEACHING Exceptional Children*, 51(4), 286–295. <https://doi.org/10.1177/0040059918808762>.
- Puga Peña, L. A., Rodríguez Orozco, J. M., & Toledo Delgado, A. M. (2016). Reflexiones sobre el lenguaje matemático y su incidencia en el aprendizaje significativo [Reflections on the mathematical language and its incidence in the significant learning]. *Sophía*, 7(20), Article 197. <https://doi.org/10.17163/soph.n20.2016.09>
- Ramírez, V., & Rosas, R. (2007). Estandarización del WISC-III en Chile: Descripción del test, estructura factorial y consistencia interna de las escalas [Standardization of the WISC-III in Chile: Test description, factor structure and internal consistency of the scales]. *Psykhe*, 16(1), 91–109. <https://doi.org/10.4067/s0718-22820070001000008>
- Rocha Carpiuc, C. (2016). Women and diversity in Latin American political science. *European Political Science*, 15(4), 457–475. <https://doi.org/10.1057/s41304-016-0077-4>
- Rosas, R. (1990). *Replikation des Berliner Intelligenzstrukturmodells(BIS) und Vorhersagbarkeit des Studienerfolgs bei chilenischen Studenten* [Replication of the Berlin Intelligence Structure Model(BIS) and predictability of academic success in Chilean students]. Freie Universität Berlin.
- Rosas, R., Espinoza, V., Martínez, C., & Santa-Cruz, C. (2022). Playful testing of executive functions with yellow-red: Tablet-based battery for children between 6 and 11. *Journal of Intelligence*, 10(4), Article 125. <https://doi.org/10.3390/jintelligence10040125>
- Rosas, R., Pizarro, M., Grez, O., Navarro, V., Tapia, D., Arancibia, S., Muñoz-Quezada, M. T., Lucero, B., Pérez-Salas, C. P., Oliva, K., Vizcarra, B., Rodríguez-Cancino, M., & von Fredeen, P. (2022). Estandarización Chilena de la Escala Wechsler de Inteligencia para Niños - Quinta Edición [Chilean Standardization of the Wechsler Intelligence Scale for Children – Fifth Edition]. *Psykhe (Santiago)*, 31(1), 1–23. <https://doi.org/10.7764/psykhe.2020.21793>
- Rosas, R., Tenorio, M., Pizarro, M., Cumsille, P., Bosch, A., Arancibia, S., Carmona-Halty, M., Pérez-Salas, C., Pino, E., Vizcarra, B., & Zapata-Sepúlveda, P. (2014). Estandarización de la Escala Wechsler de Inteligencia Para Adultos-Cuarta Edición en Chile [Standardization of the Wechsler Adult Intelligence Scale – Fourth Edition in Chile]. *Psykhe (Santiago)*, 23(1), 1–18. <https://doi.org/10.7764/psykhe.23.1.529>
- Sadowsky, S., & Martínez-Gamboa, R. (2012). *Corpus Lifcach 2.0: Word Frequency List of Chilean Spanish* [Lista de Frecuencias de Palabras del Castellano de Chile]. Retrieved December 15, 2022 from <https://zenodo.org/records/268043>
- Süß, H.-M., & Beauducel, A. (2005). Faceted models of intelligence. In O. Wilhelm, & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 313–332). Sage Publications.
- von Fritz, A., Ehlert, A., Ricken, G., & Balzer, L. (2017). *Mathematik- und Rechenkonzepte bei Kindern der ersten Klassenstufe – Diagnose*. [Mathematics and arithmetic concepts for children in first grade – diagnosis]. Hogrefe.
- von Hagen, A. (2018). *Individual differences in foreign language attainment of children with poor literacy skills* [Doctoral thesis, Macquarie University].
- von Lenhard, W., Lenhard, A., & Schneider, W. (2017). *ELFE II – Ein Leseverständnistest für Erst- bis Siebtklässler* [ELFE II – A reading comprehension test for first to seventh graders]. Hogrefe.
- Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* (4th ed.). Harcourt Assessment.
- Wechsler, D. (2014). *Wechsler Intelligence Scale for Children* (5th ed.). Pearson.
- Wright, A., & Diamond, A. (2014). An effect of inhibitory load in children while keeping working memory load constant. *Frontiers in Psychology*, 5, Article 213. <https://doi.org/10.3389/fpsyg.2014.00213>

History

Received December 31, 2022
 Revision received June 4, 2023
 Accepted November 2, 2023
 Published online February 16, 2024

Section: Methodological Topics in Assessment

Publication Ethics

All the tests on which this article is based (except the BIS) are authorized by contract between the original publishers (Pearson, Hogrefe) and the Universidad Católica de Chile. BIS was adapted as part of the first authors doctoral thesis, and there is no commercial version in Spanish.




Authorship

Ricardo Rosas, formulation and writing of the initial draft, creator of the idea of the Decalogue, writing of the introduction and three sections of the decalogue, Marcelo Pizarro, writing of five sections of the decalogue, generation of bibliography. Victoria Espinoza, writing two sections of the decalogue, generation of bibliography.

Funding

This research was supported by the ANID PIE CIE 160007 and CEDETi UC.

ORCID

Ricardo Rosas
 <https://orcid.org/0000-0002-3091-4044>
 Marcelo Pizarro
 <https://orcid.org/0000-0003-2333-804X>
 Victoria Espinoza
 <https://orcid.org/0000-0001-9462-1139>

Ricardo Rosas

Escuela de Psicología
 Pontificia Universidad Católica de Chile
 Av. Libertador Bernardo O'Higgins 340
 8331150
 Santiago
 Región Metropolitana Chile
 rrosas@uc.cl