

**Adapting a cognitive test for a different culture:
An illustration of qualitative procedures**

MAIKE MALDA¹, FONS J. R. VAN DE VIJVER², KRISHNAMACHARI SRINIVASAN³,
CATHERINE TRANSLER⁴, PRATHIMA SUKUMAR³ & KIRTHI RAO³

Abstract

We describe and apply a judgmental (qualitative) procedure for cognitive test adaptations. The procedure consists of iterations of translating, piloting, and modifying the instrument. We distinguish five types of adaptations for cognitive instruments, based on the underlying source (construct, language, culture, theory, and familiarity, respectively). The proposed procedure is applied to adapt the Kaufman Assessment Battery for Children, second edition (KABC-II) for 6 to 10 year-old Kannada-speaking children of low socioeconomic status in Bangalore, India. Each subtest needed extensive adaptations, illustrating that the transfer of Western cognitive instruments to a non-Westernized context requires a careful analysis of their appropriateness. Adaptations of test instructions, item content of both verbal and non-verbal tests, and item order were needed. It is concluded that the qualitative approach adopted here was found adequate to identify various problems with the application of the KABC-II in our sample which would have remained unnoticed with a straightforward translation of the original instrument.

Key words: Kaufman Assessment Battery for Children, Cognitive Test, Adaptation, Bias, Culture

¹ Maike Malda, Department of Psychology, Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands; email: m.malda@uvt.nl

² Tilburg University, the Netherlands and North-West University, South Africa

³ St. John's Research Institute, India

⁴ Unilever Food and Health Research Institute, the Netherlands

...You cannot take a person who for years has been hobbled by chains, bring him up to the starting line of a race and say – you are free to compete with us – and truly believe that you are treating him fairly.

Lyndon Johnson (as cited in De Beer, 2000, p. 1)

Varying definitions of fairness have been proposed; fairness can be seen as a lack of bias, as equitable treatment in a testing procedure, as equality in outcomes of testing, or as equality in opportunities to learn (American Educational Research Association, American Psychological Association, & National Council on Measurement of Education, 1999). The quote by Lyndon Johnson refers to the last definition, whereas we mainly focus on the first. Like it is unfair to run a race against a person hobbled by chains, it is unfair to assess intelligence of children from rural Africa with a test that has been validated in a Western culture (usually in the U.S. or Western Europe), with a population of children exposed to very different educational and material environments at home and school. Many children in developing and emerging countries live in multiple-risk environments and show suboptimal (physical, cognitive, and social-emotional) developmental outcomes, due to poor nutrition, housing, and hygiene, low socioeconomic status, crowded homes and classrooms, and few learning materials and opportunities (McLoyd, 1998; Walker et al., 2007). Cognitive tests of Western origin may be inadequate to assess these children; the cross-cultural suitability of these tests cannot be assumed, is often questionable, and is infrequently studied (Misra, Sahoo, & Puhhan, 1997). Since cognitive test scores are known to predict school performance of children (also in non-Westernized countries), it is important for them to be (culturally) appropriate. We propose and illustrate a systematic approach for adapting cognitive instruments to increase their cultural suitability for the target context.

Children in non-Westernized countries might be unfamiliar with testing procedures and materials, which is in sharp contrast with the relatively high level of testwiseness of Western children. For example, working with figures and puzzles may be a novel experience for children in a non-Westernized setting, whereas many Western children are exposed to these tasks from a preschool level. Making puzzles or comparable tasks can positively contribute to one's visual processing ability. Demetriou et al. (2005) found that Chinese children outperformed Greek children on tasks involving visuo-spatial processing, which the authors attributed to the massive visuo-spatial practice received in learning to write Chinese.

The use of an unsuitable instrument can lead to a biased (unfair) assessment of cognitive performance; therefore, two types of procedures have been described to reduce this bias: a priori procedures (also called judgmental procedures) and a posteriori procedures (statistical procedures). A priori procedures are applied before the instrument is administered; we refer here to all those procedures that use judgmental evidence to examine the cultural suitability of translations and adaptations of instruments, such as quality checks of translations, examinations of the adequacy of pictorial stimuli, and pilot studies to determine whether test instructions and items are interpreted as intended. A posteriori procedures are applied to the data obtained with the instrument; these involve the use of statistical methods to identify and reduce the bias in collected data (Van de Vijver & Leung, 1997). A posteriori procedures are widely used to examine differential item functioning and structural equivalence (see Ellis, 1989; Sireci & Allalouf, 2003; Sireci, Yang, Harter, & Ehrlich, 2006). We describe and apply a priori procedures in this article because their impact can be easily underrated. A

priori procedures are very relevant; problems of poor test adaptations cannot be overcome by statistical (post hoc) analyses, whatever their sophistication.

Many guidelines for test adaptations have been proposed (American Educational Research Association et al., 1999; Hambleton, 2001, 2005); yet, there is no agreement about minimum standards or best practices and very few applications have been published (e.g., Abubakar et al., 2007; Holding et al., 2004). Whereas these applications are mainly described from a procedural point of view, we conceptualize our approach by applying a systematic procedure for adapting cognitive instruments within a framework of adaptation types. We illustrate this approach by describing the adaptation process of the Kaufman Assessment Battery for Children, second edition (KABC-II) for use among 6 to 10 year-old Kannada-speaking children of low socioeconomic status in Bangalore, India. Our aim was to develop a measure of children's cognitive performance that is suitable for this particular context and to learn lessons from this adaptation procedure that could generalize to other settings and cognitive test batteries, such as the Wechsler scales (Wechsler, 1949, 1974, 1991, 1997, 2004) or the newly developed Adaptive Intelligence Diagnosticum (Kubinger, 2004; Kubinger, Litzenberger, & Mrakotsky, 2007; Kubinger & Wurst, 2000).

Test adaptation procedure

The adaptation procedure that is proposed and illustrated here has two core elements. The first refers to how the procedure is conducted. Our procedure consists of an iterative process of implementing modifications to an instrument and using judgmental evidence to examine the adequacy of the modifications. This procedure is in line with what is called "cognitive pretesting" or "cognitive interviewing" (DeMaio & Rothgeb, 1996; Willis, 2005), which refers to a method to evaluate whether the target audience properly understands, processes, and responds to the test items. Cognitive pretesting uses think-aloud and verbal probing procedures, and has been mainly applied to evaluate surveys; yet, it can be used to test any type of test material. A criterion for the success of a judgmental procedure such as cognitive pretesting is that all items of the battery are interpreted as intended. The second core element of our procedure refers to which types of adaptation are involved; a taxonomy of adaptation types is proposed here that can be used in any adaptation procedure. Before presenting the taxonomy we describe the various kinds of bias that may need to be accounted for in test adaptations.

Bias in testing

In cross-cultural research, bias is a generic term for all kinds of factors that threaten the validity of intergroup comparisons (Van de Vijver & Hambleton, 1996). Bias is a consequence of a test's cultural loading, which refers to the extent to which the test implicitly or explicitly refers to a particular cultural context. There are three main types of bias: *construct bias*, *method bias*, and *item bias* (for a detailed description see Van de Vijver & Poortinga, 2005, and Van de Vijver & Tanzer, 2004). An instrument that shows construct bias in a cross-cultural comparison does not measure the same psychological concept across cultures. We did not focus on construct bias in our adaptation because we focus here on intelligence in

a school context and the underlying structures of many cognitive test batteries presumably are universally applicable (Berry, Poortinga, Segall, & Dasen, 2002; Georgas, Weiss, Van de Vijver, & Saklofske, 2003; Irvine, 1979; Van de Vijver, 1997). Method bias refers to sources of bias that arise from methodological aspects of a study, such as instrument bias and administration bias. Item bias (differential item functioning) refers to item-specific problems in cross-cultural comparisons, such as item ambiguity due to poor item translations or culture-specific elements (e.g., an item about a vacuum cleaner is biased against cultures in which this appliance is uncommon). The described forms of bias can be remedied by adaptation.

Adaptation, adoption, and assembly

Adaptation is a way to maximize the cultural appropriateness of an instrument and thereby to minimize bias. *Adaptation* has become the generic term for any procedure in which an instrument that is developed for one cultural group is transferred for usage in another cultural group. The term has replaced the traditional concept of *translation*, because of the growing appreciation that transferring a test to a new cultural and linguistic context involves more than merely translating an instrument (producing a linguistically equivalent version in another language).

The term adaptation is also used in a more specific sense. Three terms have been proposed to describe the transformations that may be needed to transfer an instrument to another culture: *adoption* (or *application*), *adaptation*, and *assembly* (Hambleton & Patsula, 1998, 1999; Van de Vijver, 2003; Van de Vijver & Poortinga, 2005). Adoption of an instrument comes down to a close translation into the target language, and can be used if the purpose of a study is to compare scores across cultures directly (Van de Vijver, 2003). Assembly involves the construction of an entirely new instrument, and is usually applied when the translation of an existing instrument would yield an entirely inappropriate measure in the target culture or when the study concerns a new research topic for which no suitable instrument is available yet (Harkness, Van de Vijver, & Johnson, 2003). Adaptation has features of both adoption and assembly; it amounts to a combination of close translation of the parts of the instrument that are assumed to be adequate in the target culture, such as test instructions and items, and a change of other parts when a close translation would be inadequate for linguistic, cultural, or psychometric reasons (Hambleton & De Jong, 2003; Harkness, Mohler, & Van de Vijver, 2003).

The two different usages of the term adaptation (broad and specific) are fairly compatible if we do not see adoption, adaptation, and assembly as three entirely different kinds of procedures, but as labels on a continuum that ranges from a close translation of all instrument features (adoption) to a complete change of these features (assembly). Adaptation can then be seen as a term for all transfers that do not belong to the extremes of the continuum. In this interpretation, adaptation covers a wide range of changes to tests (which may explain the popularity of adaptation in the current literature) and is the main method of transfer in our current qualitative evaluation of test appropriateness.

Types of adaptation

Adaptations can amount to various types of changes (Harkness, Van de Vijver et al., 2003; Van de Vijver, 2006). We propose a framework of types of adaptation which can help us to systematize the adaptation process and the choices made in this process. In our view, five types can be distinguished that are relevant in the context of adapting cognitive tests. *Construct-driven adaptations* are related to differences in definitions of psychological concepts across cultures (e.g., when the aim is to measure “intelligence”, the test should be adapted according to the target culture’s definition of intelligence). *Language-driven adaptations* result from the unavailability of semantically equivalent words across languages (e.g., there is no Dutch equivalent for the English word “distress”) or from structural differences between languages (e.g., words or grammatical structures automatically refer to gender in some languages, which makes it difficult to avoid gender-specific references. For example, the English word “friend” can indicate both a male and a female person, whereas the German word “Freund” refers to a male friend and “Freundin” to a female friend). *Culture-driven adaptations* result from different cultural norms, values, communication styles, customs, or practices (e.g., an item about the celebration of birthdays should take into account that cultures differ considerably in practices and cultural relevance of birthdays). *Theory-driven adaptations* involve changes that are required because of theoretical reasons (e.g., digit span items should ideally have digit names that are all of similar length. Similarity in digit length may be lost when the items are translated into another language). The last type are *familiarity/recognizability-driven adaptations* which are based on differential familiarity with task or item characteristics (e.g., a prototypical drawing of a house in one culture is not necessarily recognized as such in another culture) or stimulus materials (e.g., in some cultures children might not be used to manipulate geometric shapes). Different types of adaptations are applicable to different types of tests. We consider these five types of adaptations sufficient to describe the changes that are required in making cognitive instruments suitable for new cultural contexts. The framework introduced here is used to indicate which adaptations we have used to improve the cultural suitability of the KABC-II for our Indian sample and to place our findings into the broader perspective of adapting cognitive tests in general.

Adapting the Kaufman Assessment Battery for Children, second edition

The KABC-II (a revised and re-standardized second edition of the K-ABC) is an individually administered measure of cognitive ability that can be used for children from 3 to 18 years of age (Kaufman & Kaufman, 2004) and measures short-term memory, visual processing, long-term storage and retrieval, fluid reasoning, and crystallized abilities. The test combines three characteristics that make it promising for research and applications in non-Westernized countries: (1) the KABC-II is based on a theoretical model (the Cattell-Horn-Carroll model of broad and narrow abilities; Carroll, 1993; McGrew, 2005) that is assumed to have a universal validity; (2) the test has been designed to minimize the influence of language and cultural knowledge on test results; (3) the test contains teaching items, that ensure understanding of the task demands.

The present study is relevant in providing information about the (in)appropriateness of the KABC-II among Kannada speaking children in Bangalore. Furthermore, the relevance of

our qualitative adaptation procedure goes beyond the immediate context of the present instrument and cultural context for two reasons. Firstly, the instrument shows generalizability to other, widely used cognitive batteries regarding instruction, item, and response formats. Secondly, the adaptation deals with large cultural, linguistic, and socioeconomic differences between the original Western (American) context and the non-Westernized target (Indian) context. The larger these differences, the more salient the (possible) bias, providing good conditions for a critical test of why and for which test aspects adaptations are required. Many other cross-cultural studies on the application of cognitive tests (such as the WISC-III by Georgas et al., 2003) do not include samples that differ substantially from the original test sample in cultural or educational background.

Method

Participants

Our adaptation is part of a larger study among children of low socioeconomic status in Bangalore (state of Karnataka, South India). Fifty seven Kannada-speaking children took part in the adaptation process (31 boys and 26 girls), they were between 6 and 10 years old ($M = 8.08$) and from grades one to five from five primary schools. The number of children participating in our adaptation could not be determined nor accurately estimated beforehand, because in each step of the iterative procedure of translating, piloting (i.e., cognitive pre-testing) and modifying that we employed, a new (small) sample of children was involved and for each individual subtest the iterations continued until the adaptations were deemed satisfactory (see Procedure). As a consequence, the number of children involved in the pilot testing differed across the subtests.

Context

Information about the children's direct living environment, needed for an adequate adaptation, was collected by visiting homes and schools and interviewing parents and teachers. We wanted to learn what type of cognitive stimulation was provided to the children by their environment. There were very few or no toys to play with and usually no other learning materials than school books were present in the homes. Most families owned a television. Children either played outside in the streets or watched television when not doing chores. Interviews with teachers revealed that rote learning is a commonly applied teaching technique. This technique is well applicable with large numbers of children and with a collectivistic style of teaching, where children are hardly addressed individually.

Procedure

In line with practices recommended in the literature on adaptation guidelines (e.g., Geisinger, 1994; Hambleton, 2005; Hambleton & Patsula, 1999), we employed an iterative procedure of translating, piloting, and modifying instructions, examples and items if needed.

The adaptation process took eight months from developing the initial ideas to completing the final test battery.

A team of four psychologists (all fluent in both Kannada and English, and with a Master's degree in Psychology, specialized in Child Psychology) translated the test instructions and items of the KABC-II from (American) English into Kannada. We instructed the team to try to avoid poor readability and lack of naturalness, which are well known problems of close translations (Harkness, 2003; Stansfield, 2003). The translation was independently back translated by a psychologist. The translated version was fine-tuned during the pilot test through iterations of modifying translations, administering these modifications to other children of the pilot sample, and implementing further modifications, if needed. Some subtests required more extensive piloting than others and each new subtest version was administered in a new round of piloting to a different set of children so as to avoid learning effects from previous test versions. The iterative process was continued until the subtest version was found to be adequate (i.e., the children showed understanding of the instructions and concepts by performing well on at least the first few items). The adapted instruments are described in more detail in the Results section.

The test administration in our pilot test was done in a non-standard way (Van de Vijver & Tanzer, 2004) in order to evaluate the appropriateness of the test materials and test procedure. In this non-standard way, the focus is not primarily on the child's responses to test items, but on identifying the processes behind these responses. One test examiner (a trained psychologist) administered KABC-II subtests to all children in the pilot. A supervising psychologist (first author) observed each of these test administrations. The examiner asked the child to repeat the instructions when there was any doubt about whether a child had understood the instructions of a particular subtest. The child was asked to explain his/her answer if an answer had to be selected from various options. Both the supervisor and the test examiner evaluated the child's ability to work with the test materials and the response formats. This supervisor also assessed the skills of the test examiner to administer the various adapted subtests. The extensive practice ensured that the examiner administered the items in an appropriate way, as described by Kaufman and Kaufman (2004) so that administration bias could be minimized.

Results

We focused on eight of the core subtests of the KABC-II for 7-12 year-old children. The Results section is divided in three parts. The subtests that required a theory-driven adaptation are presented first (Number Recall and Atlantis), followed by the subtests that required a familiarity/recognizability-driven adaptation (Triangles, Rover, Pattern Reasoning, and Story Completion), and finally the subtests that required both types of adaptation (Word Order and Rebus). Each subtest is first described (Kaufman & Kaufman, 2004), followed by an overview of the main modifications. Only those aspects of the adaptation process are described here that we expect to be relevant for adaptations of other cognitive tests.

Theory-driven adaptations

Number Recall (short-term memory). In this task, the child is asked to repeat a series of monosyllabic digits (1 to 9, excluding 7) in the same sequence as presented by the examiner, with series ranging in length from two to nine digits. Number Recall is comparable to Digit Span (forward) from the Wechsler scales and to many other short-term memory tests in various cognitive test batteries, such as Immediately Reproducing from the AID 2 (Kubinger & Wurst, 2000). According to Baddeley's phonological loop model (Baddeley, Thomson, & Buchanan, 1975; Cowan, Baddeley, Elliott, & Norris, 2003), the number of items that can be stored in memory varies with their phonological length (such as the number of syllables). The shorter the items, the more items can be recalled. It follows from the model that Number Recall will be more sensitive to differences in memory capacity when shorter digits are used and that it is important to maintain a constant phonological digit length.

All digits in Kannada from 1 to 9 are bisyllabic, except 2 and 9, which have three syllables. We decided to rely as much as possible on the bisyllabic digits in the Kannada version. The three-syllabic digits (2 and 9) were only introduced late in the test, in series of eight and nine digits.

Atlantis (long-term storage and retrieval). The examiner teaches the child nonsense names (here defined as pseudo-words that have a common phonological structure) for fanciful pictures of fish, plants, and shells. The child has to point to the corresponding picture in an array of pictures when it is named. The test measures the ability to memorize new phonological information without the support of the meaning or context of the words. A comparable task is Memory for Names from the Woodcock-Johnson III tests of cognitive ability (Woodcock, McGrew, & Mather, 2001). For the use of nonsense syllables, see also the AID 2 (Kubinger & Wurst, 2000).

The first group of Kannada children (who are not familiar with the English language) found it difficult to make distinctions between the English nonsense names. Therefore, we replaced the English nonsense names by Kannada nonsense names. The sounds of the chosen names were sufficiently distinct for the children to easily distinguish between the words. As in the original version, one-, two-, and three-syllable names were chosen for fish, plants, and shells, respectively.

Familiarity/recognizability-driven adaptations

Triangles (visual processing). The child assembles several identical foam triangles (blue on one side and yellow on the other) to match a target picture of an abstract design. For easier items, the child assembles a set of colorful plastic shapes to match a model constructed by the examiner or shown in the test booklet. The test is based on Koh's (1927) Block-Design Test and shows similarities with subtests such as Block Design from the Wechsler scales, Pattern Construction from the Differential Ability Scales (Elliott, 1990), and Analyzing and Synthesizing from the AID 2 (Kubinger & Wurst, 2000).

Subsequent items of Triangles should increase in difficulty, as is the case for all KABC-II subtests and for most subtests of other cognitive test batteries. It became clear during the pilot test that compliance with this rule required changes in the order and nature of some items. The original sample item of the foam triangles involves constructing a larger triangle

with two smaller ones. This item appeared to be too difficult for a sample item. Furthermore, the children in the pilot test could solve items relatively well when the triangles in the target figure showed left-right symmetry, but items without this left-right symmetry were much more difficult for them. This could be the result of their lack of experience with making puzzles. We decided to include three items with one triangle in the adapted test so that children could explore the possibilities of manipulating a single triangle before they had to manage two or more. We also added one easier two-triangle item and slightly changed the item order to ensure an increasing level of difficulty for the Kannada children.

The original test manual indicates that for most items any rotation of the final (total) configuration should be scored as correct. The pilot test showed that children sometimes produced solutions with a large rotation relative to the target figure, which would have to be scored as correct. However, when the children were asked to explain their solution, they did not show full understanding of the item. To avoid this problem, we decided that only solutions with a rotation of 45 degrees or less in either direction from the displayed model would be scored as correct.

Part of the Triangles test is timed. Because the local schools do not train their children in managing their time and performing quickly while doing exercises or tests, we decided to apply a more liberal time limit: the original time limits were relaxed by 15 seconds. No extra points were given for quick responses, having only 0 (incorrect) and 1 (correct) as possible scores.

Rover (visual processing). The child has to move a dog toy (called Rover) to a bone on a checkerboard-like grid that contains obstacles (rocks and weeds) by making as few moves as possible. Rover is based on several non-verbal problem-solving tasks, such as the Tower of Hanoi (Cook, 1937).

When the original Rover dog was used to make the moves, the children tended to start the path to the bone in the direction the dog was facing. To prevent this, we needed an object that is similar on all sides so that it does not implicitly suggest a direction to the child. We replaced the original dog by a pawn, which turned out to be well accepted by the children.

Not all children in the pilot test understood which moves Rover was allowed to make. To overcome this problem, we adapted one sample item and changed two regular test items into sample items to ensure that the child understood the principles of the test completely (e.g., regarding diagonal moves and regarding some obstacles drawn on the grid, like a rock). Three test items were added to give the child the opportunity to show that the principle of the test was understood before moving on to the next phase (in which a rock was introduced, which should be avoided when moving the dog to the bone). Like in Triangles, the original time limits were relaxed by 15 seconds.

Pattern Reasoning (fluid reasoning). The child is shown a series of stimuli that form a logical sequence organized according to a pattern that is not explicitly provided (e.g., A-B-A-?-A); one stimulus in the series is missing. The child completes the pattern by selecting the correct stimulus from an array of four to six options at the bottom of the page. Most stimuli are abstract, geometric shapes, and some easy items use meaningful pictures. Pattern Reasoning shows similarities with the subtest Matrix Reasoning from the WISC-IV (Wechsler, 2004) and with Raven's Standard (Raven, Raven, & Court, 1998b) and Coloured (Raven, Raven, & Court, 1998a) Progressive Matrices.

Two adaptations were required. Firstly, we slightly changed the administration of the second item (a teaching item) where children often appeared to choose the correct answer

option without understanding the pattern. Some children indicated that this was because the correct option is an appealing picture. It was therefore decided to explain the correct answer regardless of whether the child's answer was incorrect or correct. Secondly, the original version requires the assessment of response times at item level. We did not monitor time because the pilot test showed that accurate measures of the short response times (often only a few seconds) were difficult to obtain, leaving only 0 (incorrect) and 1 (correct) as possible scores.

Story Completion (fluid reasoning). The child is shown a row of pictures that tell a story, but some of the pictures are missing. The child is given a set of pictures, selects the ones that are needed to complete the story, and places the missing pictures in their correct locations.

The subtest contains many references to cultural aspects that were unfamiliar or unknown to our target population (in general or because of their low socioeconomic status). Examples are having a birthday party, blowing balloons, specific Western dishes, and the use of napkins. We replaced the entire subtest (culture-driven adaptation) by our items based on the items of Picture Arrangement from the Wechsler Intelligence Scale for Children (Wechsler, 1949, 1974, 1991), which shows similarities with Social and Material Sequencing from the AID 2 (Kubinger & Wurst, 2000). Each item of Picture Arrangement consists of a series of pictures depicting a story. The pictures are presented in an incorrect order and the child is asked to arrange them in an order that makes a sensible story. Although Picture Arrangement seemed to be less related to a specific cultural context, the items needed modification.

The WISC Picture Arrangement (Wechsler, 1949) and the WISC-R Picture Arrangement (Wechsler, 1974) were each administered to approximately 10 children to get a basic idea of test aspects that should be adapted. The findings, combined with extensive discussions with the local study team, were our starting point for developing the adapted version. New drawings and modifications in drawings were made by a local artist. All items were extensively piloted. The number of cards in each item was kept similar to the original Wechsler scales whenever possible. Five new themes were introduced (two sample items and three test items), one item from the original WISC was used as well as one item from WISC-III, and eight items of WISC-R were adapted.

There is only one sample item in the original Picture Arrangement task; furthermore, the item does not require any active participation of the child. The examiner arranges the cards in the correct order, tells the story, and asks the child whether he or she understood the item. We decided to include two sample items that require active participation of the child. The administrator first puts the cards in the correct order and tells the displayed story; the administrator then puts the cards in the incorrect order again and asks the child to arrange them in the correct order. The child then has to point to each card and tell the story depicted. The administrator explains the item further (again) if needed, until the child has clearly understood the item.

Stories with a high cultural loading were removed (i.e., items that the children could not understand because the concepts expressed or objects displayed in the items were not familiar or recognized), items with a lower cultural loading were adapted, and some new items were created. The sample item of both the WISC and the WISC-R is a three-card item that shows how a lady walks to a scale, takes her weight, and walks away. We decided to remove the item because the type of scale that is used in the item is unfamiliar to the children in our target sample. An example of an adapted item is a four-card item describing a burglar break-

ing into a house and getting caught by the police. The pilot test made clear that Kannada children did not recognize the cues in the outfit of the burglar (horizontally black-and-white striped shirt in combination with a small mask over the eyes). In addition, children are not familiar with windows that slide vertically. In the adapted version, the burglar has an Indian appearance and the window has two glass panes that open sideways (see Figure 1).



Figure 1:

Example of a culturally adapted drawing of Picture Arrangement

Theory-driven and familiarity/recognizability-driven adaptations

Word Order (short-term memory). The child has to point to a series of silhouettes of common objects in the same order as the examiner said the names of the objects while they were out of the child's sight; an interference task (color naming) is added between the stimulus and the response for the more difficult items. Stimuli of the American version of Word Order were selected carefully to ensure that young children with normal language development would readily identify and label all pictures in an adequate manner. The American original contains only objects with monosyllabic names to control phonological length and complexity similarly to what was previously observed for Number Recall (theory-driven adaptation). The test is based on auditory-vocal short-term memory tests, in which the child has to repeat a series of unrelated words spoken by the test examiner. Word Order is different from these traditional tests in that it does not require a verbal response from the child.

Everyday objects with monosyllabic names in Kannada were difficult to find, which made it necessary to select everyday objects with bisyllabic names (theory-driven adaptation). The additional criteria for choosing new stimuli were that their names and corresponding visual representation (black-and-white drawings) should be unambiguous and highly familiar (familiarity/recognizability-driven adaptation). One out of the twelve original stimuli needed redrawing; the drawing of a house contained a chimney, which was not known to

the Indian children and was therefore removed. Six out of the twelve original stimuli needed replacement. Drawings of a star, key, hand, moon, heart, and shoe were replaced by drawings of a flower, book, leg, sun, chair, and bus, respectively. The goal of the color interference task (color naming) is to measure recall following interference. Children had problems with naming gray blocks because there is no common Kannada word for gray. This problem was avoided by using blocks with more familiar colors.

Rebus (long-term storage and retrieval). In this test measuring associative memory, (verbal) learning, and long-term storage and retrieval, the examiner teaches the child the word or concept associated with each particular drawing, and the child “reads” aloud phrases and sentences composed of these drawings (e.g., six different drawings can form the sentence “The girl and boy play games”). A comparable test is Visual-Auditory Learning from the Woodcock-Johnson III (Woodcock et al., 2001). We did not administer Rebus. Translating and adapting would have been very difficult in Kannada language. The sentences to be produced are so strongly related to the specifics of the local language (such as the use of particles and word order in a sentence), that a close (literal) translation was not possible and a modification would produce a version that is considerably different from the original.

We replaced Rebus by our Verbal Learning Test that is based on the Rey Auditory Verbal Learning Test (Rey, 1964) (language-driven adaptation). The Verbal Learning Test measures immediate memory, efficiency of learning, and recall after short and long delay periods. Although the nature of this test differs from Rebus in that it does not associate verbal labels with visual stimuli (associative memory), both tests focus on storing and efficiently retrieving newly learned information.

Our test consists of a list of 15 words. The following criteria were used for choosing words in the list: (a) the words are related to children’s everyday experience, which ensures high familiarity; (b) the words belong to the same grammatical category (e.g., nouns) and refer to concrete objects; (c) the words have two syllables; (d) phonological similarities between words in the list are kept to a minimum; (e) the words do not belong to the same semantic category (e.g., animals or means of transport) in order to prevent clustered recall; (f) the words are not used elsewhere in the cognitive test battery. Criterion (a) refers to familiarity/recognizability-driven adaptations, whereas criteria (b) to (f) illustrate theory-driven adaptations. The list is read out loud to the child at a rate of one word per second and at a constant tone. Then the child is asked to reproduce all the words from the list that can be remembered. This procedure is repeated twice and after a 20 minute delay during which two other cognitive tests are administered, recall is measured for the fourth time.

Discussion

Many cognitive tests have been developed in the United States and Europe. If these tests are used in a non-Westernized context, various adaptations (involving instructions, item formats, response formats, and test stimuli) may be needed to ensure their suitability for the new cultural context. Our focus has been entirely on judgmental, a priori procedures of the test adaptation process; we did not address the adaptation from an a posteriori, statistical point of view. Because no agreement exists on minimum standards or best practices for judgmental procedures, we proposed and applied a systematic, qualitative approach to adapt cognitive tests. Our approach combines two aspects. Firstly, we systematically employed

iterations of translating, piloting (i.e., cognitive pretesting), and modifying items. Secondly, we based the adaptations on a taxonomy of types of cognitive test adaptations we presented. Our approach is illustrated by an adaptation of the Kaufman Assessment Battery for Children, second edition (KABC-II) for use among 6 to 10 year-old Kannada-speaking children of low socioeconomic status in Bangalore, India. The adaptation dealt with cultural, linguistic, and socioeconomic differences between the original (American) context and the target (Indian) context. Our procedure and findings provide us with valuable information that can be generalized to the cross-cultural use of other cognitive tests (such as the Wechsler scales or the AID 2) and other settings.

Adaptations of all subtests were needed to maximize the suitability of the (American) KABC-II for use in our Indian sample because many subtests showed implicit or explicit references to cultural elements. Theory-driven adaptations were applied in Number Recall and in Atlantis. Familiarity/recognizability-driven adaptations were used in Triangles, Rover, Pattern Reasoning, and Picture Arrangement. In Word Order and Verbal Learning Test, both familiarity/recognizability-driven adaptations and theory-driven adaptations were applied. We can conclude that most adaptations were needed because of problems with the familiarity and recognizability of specific tasks (e.g., the subtest Rover) and of specific items (e.g., the drawing of a key in the American Word Order). A translation of the test without the adaptations is presumably highly susceptible to instrument bias (i.e., a form of method bias) and item bias; an inadequately adapted instrument is likely to provide an underestimation of the cognitive performance of a child.

We introduced a distinction between five types of adaptations that can be used in transferring instruments to a new linguistic/cultural context. This categorization allows us to draw conclusions about our KABC-II adaptation and about cognitive test adaptations in general. Firstly, two types of adaptation were sufficient to reduce the cultural unsuitability of the eight selected subtests. The nature of the test clearly determines the types of adaptation needed. For instance, language-driven adaptations may be more relevant for questionnaires or for predominantly verbal cognitive tests (e.g., WISC subtests like Vocabulary and Similarities) that measure crystallized abilities. Some core KABC-II subtests measure these abilities (Riddles and Verbal Knowledge); however, we did not include these because of their presumed high cultural loading. Culture-driven adaptations may be more relevant for subtests such as the Comprehension (WISC), in which questions are asked that refer to social situations and conventions.

Secondly, familiarity/recognizability-driven adaptations were more laborious than theory-driven adaptations; the former assume thorough cultural knowledge (local people were our cultural informants), these adaptations can often take many forms and require a choice out of many candidate solutions, and these adaptations require elaborate piloting to evaluate the success of (each successive version of) the adaptation. Theory-driven adaptations, on the other hand, are more straightforward and less susceptible to disagreement, because the underlying principles are widely investigated and documented. As a result, smaller pilot samples and fewer iterations (only one or two) were needed to modify subtests that required theory-driven adaptations (such as Number Recall) than the familiarity/recognizability-driven adaptations (at least four or five) before an acceptable level of linguistic/cultural suitability was reached. An additional reason for the relative ease of performing theory-driven adaptations is that the abilities measured by those subtests (memory and learning) are

very familiar to children that are frequently addressed by a teaching technique based on rote learning.

What are the implications of our adaptation procedure for the use and adaptation of other instruments in a non-Westernized context? Firstly, many adaptations were needed for the KABC-II, indicating the necessity to closely inspect all Western instruments that are to be used or were already used outside their culture of origin for possible sources of bias. Secondly, some of our adaptations were more general and would presumably apply to various non-Westernized contexts whereas other adaptations seem to be more culture-specific. The addition of test instructions and items to ensure children's understanding of the (sub)test concept seems to be universally relevant (and especially relevant for children without assessment experience). On the other hand, the results of theory-driven and familiarity/recognizability-driven adaptations are specific for a particular culture and may therefore not be universally applicable. Thirdly, we would like to stress the importance of paying attention to the cultural loading of tests with non-verbal stimuli, in particular when there are large differences between the cultures of the test developer and the participants. As opposed to verbal tests with culture-related stimuli (e.g., reading tasks, spelling tasks, the WISC subtest Comprehension), tests with non-verbal stimuli are often considered to travel well across cultures due to their limited emphasis on language (Ortiz & Dynda, 2005); however, non-verbal tests are not "culture-free" (cf. Helms-Lorenz, Van de Vijver, & Poortinga, 2003). Fourthly, familiarity/recognizability driven adaptations do not merely entail changes in the content of the items; they can also focus on response formats (e.g., children in some contexts are not used to working with multiple choice response formats), and on the order in which items are presented if an increase in item difficulty is required. Finally, our study points to the crucial importance of combining various fields of expertise in the adaptation process. Linguistic, psychometric, and cultural knowledge should be combined to successfully adapt an instrument; in the case of this particular adaptation, knowledge on intelligence theories and child psychology were combined with linguistic and cultural expertise. We would specifically like to emphasize the need to work with cultural informants. Our adaptation involved local study collaborators (some had an expertise in psychology, others were experts in the local language) as well as the people who were most directly involved with children in our target population, such as parents (to provide information on the child's cognitive stimulation at home) and teachers (to provide information on the school curricula and teaching strategies). An adequate test adaptation requires extensive observations of the children's natural home and school environment, including child raising and teaching methods.

Our focus has been entirely on a priori procedures of the test adaptation process. Obviously, studies of the adequacy (validity) of adaptations should be complemented by statistical, a posteriori evidence (through data collection and data analysis). After data have been collected with an adapted instrument, various statistical procedures need to be employed to examine to what extent the original goals of developing an appropriate test have been accomplished. For example, the question has to be addressed whether the expected factor structure can be found, whether the adapted subtests constitute reliable and bias-free measures. In short, the data collection provides the litmus test of the adequacy of the adaptation (Malda, Van de Vijver, Srinivasan, Transler, & Sukumar, 2008). An elaborate, detailed, and systematic test adaptation in our view constitutes a first, important, and strongly recommended step in assessing cognitive abilities with any Western (cognitive) test in a non-Westernized context.

Acknowledgements

We would like to thank Unilever Food and Health Research Institute (Vlaardingen, the Netherlands) for sponsoring the study. Our gratitude goes to Kamala, Amitha, Sapna, and Carol for their work on the translations and their input in the adaptations. We would like to thank Dr. Ashok (Bangalore University) for comments on the translations and GDU (Bangalore) for the artwork they prepared for some of the subtests. We are very grateful to Sumithra Muthayya and Ans Eilander for their indispensable support.

References

- Abubakar, A., Van de Vijver, F. J. R., Mithwani, S., Obiero, E. L., N., Kenga, S., & Katana, K. H., P. (2007). Assessing developmental outcomes in children from Kilifi, Kenya, following prophylaxis for seizures in cerebral malaria. *Journal of Health Psychology, 12*, 417-430.
- American Educational Research Association, American Psychological Association, & National Council on Measurement of Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*, 575-589.
- Berry, J. W., Poortinga, Y. H., Segall, M. H., & Dasen, P. R. (2002). *Cross-cultural psychology: Research and applications*. Cambridge, UK: Cambridge University Press.
- Cook, T. W. (1937). Amount of material and difficulty of problem solving. II. The disk transfer problem. *Journal of Experimental Psychology, 20*, 288-296.
- Cowan, N., Baddeley, A. D., Elliott, E. M., & Norris, J. (2003). List composition and the word length effect in immediate recall: A comparison of localist and globalist assumptions. *Psychonomic Bulletin & Review, 10*, 74-79.
- De Beer, M. (2000). *The construction and evaluation of a dynamic computerised adaptive test for the measurement of learning potential*. Unpublished doctoral dissertation, University of South Africa, Pretoria.
- DeMaio, T. J., & Rothgeb, J. M. (1996). Cognitive interviewing techniques: In the lab and in the field. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for cognitive and communicative processes in survey research* (pp. 177-196). San Francisco: Jossey-Bass.
- Demetriou, A., Kui, Z. X., Spanoudis, G., Christou, C., Kyriakides, L., & Platsidou, M. (2005). The architecture, dynamics, and development of mental processing: Greek, Chinese, or universal? *Intelligence, 33*, 109-141.
- Elliott, C. D. (1990). *Differential Ability Scales*. San Antonio, TX: Psychological Corporation.
- Ellis, B. B. (1989). Differential item functioning: Implications for test translations. *Journal of Applied Psychology, 74*, 912-921.
- Geisinger, K. F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment, 6*, 304-312.
- Georgas, J., Weiss, L. G., Van de Vijver, F. J. R., & Saklofske, D. H. (Eds.). (2003). *Culture and children's intelligence: Cross-cultural analysis of the WISC-III*. San Diego, CA: Academic Press.
- Hambleton, R. K. (2001). The next generation of the ITC Test Translation and Adaptation Guidelines. *European Journal of Psychological Assessment, 17*, 164-172.

- 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* (pp. 3-38). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hambleton, R. K., & De Jong, J. H. A. L. (2003). Advances in translating and adapting educational and psychological tests. *Language Testing*, *20*, 127-134.
- Hambleton, R. K., & Patsula, L. (1998). Adapting tests for use in multiple languages and cultures. *Social Indicators Research*, *45*, 153-171.
- Hambleton, R. K., & Patsula, L. (1999). Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices. *Journal of Applied Testing Technology*, *1*, 1-30.
- Harkness, J. A. (2003). Questionnaire translation. In J. A. Harkness, F. J. R. Van de Vijver, & P. P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 35-56). Hoboken, NJ: John Wiley & Sons, Inc.
- Harkness, J. A., Mohler, P. P., & Van de Vijver, F. J. R. (2003). Comparative research. In J. A. Harkness, F. J. R. Van de Vijver, & P. P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 3-16). Hoboken, NJ: John Wiley & Sons.
- Harkness, J. A., Van de Vijver, F. J. R., & Johnson, T. P. (2003). Questionnaire design in comparative research. In J. A. Harkness, F. J. R. Van de Vijver, & P. P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 19-34). Hoboken, NJ: John Wiley & Sons.
- Helms-Lorenz, M., Van de Vijver, F. J. R., & Poortinga, Y. H. (2003). Cross-cultural differences in cognitive performance and Spearman's hypothesis: *g* or *c*? *Intelligence*, *31*, 9-29.
- Holding, P. A., Taylor, H. G., Kazungu, S. D., Mkala, T., Gona, J., Mwamuye, B., et al. (2004). Assessing cognitive outcomes in a rural African population: Development of a neuropsychological battery in Kilifi District, Kenya. *Journal of the International Neuropsychological Society*, *10*, 246-260.
- Irvine, S. H. (1979). The place of factor analysis in cross-cultural methodology and its contribution to cognitive theory. In L. H. Eckensberger, W. J. Lonner, & Y. H. Poortinga (Eds.), *Cross-cultural contributions to psychology* (pp. 300-341). Lisse, the Netherlands: Swets and Zeitlinger.
- Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children, Second Edition: Manual*. Circle Pines, MN: AGS Publishing.
- Kohs, S. C. (1927). *Intelligence measurement*. New York: Macmillan.
- Kubinger, K. D. (2004). On a practitioner's need of further development of Wechsler Scales. Adaptive Intelligence Diagnosticum (AID 2). *The Spanish Journal of Psychology*, *7*, 101-111.
- Kubinger, K. D., Litzenberger, M., & Mrakotsky, C. (2007). A new perspective of traditional intelligence theories through modern test conceptualization. *Studia Psychologica*, *49*, 295-311.
- Kubinger, K. D., & Wurst, E. (2000). *Adaptives Intelligenz Diagnostikum (AID 2)*. [Adaptive intelligence diagnosticum]. Göttingen, Germany: Beltz.
- Malda, M., Van de Vijver, F. J. R., Srinivasan, K., Transler, C., & Sukumar, P. (2008). *Adapting a cognitive test for a different culture: An illustration of qualitative procedures*. Manuscript submitted for publication
- McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities: Past, present, and future. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 136-181). New York: The Guilford Press.
- McLoyd, V. C. (1998). Socioeconomic disadvantage and child development. *American Psychologist*, *53*, 185-204.

- Misra, G., Sahoo, F. M., & Puhan, B. N. (1997). Cultural bias in testing: India. *European Review of Applied Psychology*, 47, 309-316.
- Ortiz, S. O., & Dynda, A. M. (2005). Use of intelligence tests with culturally and linguistically diverse populations. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 545-556). New York: The Guilford Press.
- Raven, J., Raven, J. C., & Court, J. H. (1998a). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 2: The Coloured Progressive Matrices*. San Antonio, TX: Psychological Corporation.
- Raven, J., Raven, J. C., & Court, J. H. (1998b). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 3: The Standard Progressive Matrices*. San Antonio, TX: Psychological Corporation.
- Rey, A. (1964). *L'examen clinique en psychologie*. [The clinical examination in psychology]. Paris: Presses Universitaires de France.
- Sireci, S. G., & Allalouf, A. (2003). Appraising item equivalence across multiple languages and cultures. *Language Testing*, 20, 148-166.
- Sireci, S. G., Yang, Y., Harter, J., & Ehrlich, E. J. (2006). Evaluating guidelines for test adaptations: A methodological analysis of translation quality. *Journal of Cross-Cultural Psychology*, 37, 557-567.
- Stansfield, C. W. (2003). Test translation and adaptation in public education in the USA. *Language Testing*, 20, 189-207.
- Van de Vijver, F. J. R. (1997). Meta-analysis of cross-cultural comparisons of cognitive test performance. *Journal of Cross-Cultural Psychology*, 28, 678-709.
- Van de Vijver, F. J. R. (2003). Test adaptation/translation methods. In R. Fernández-Ballesteros (Ed.), *Encyclopedia of psychological assessment* (pp. 960-963). Thousand Oaks, CA: Sage Publications.
- Van de Vijver, F. J. R. (2006, July). *Toward the next generation of instruments in cross-cultural testing: Recent developments in translations and adaptations*. Paper presented at the ITC 5th International Conference on Psychological and Educational Test Adaptation across Language and Cultures, Brussels.
- Van de Vijver, F. J. R., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist*, 1, 89-99.
- Van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage Publications.
- 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* (pp. 39-63). Mahwah, NJ: Lawrence Erlbaum Associates.
- Van de Vijver, F. J. R., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology*, 54, 119-135.
- Walker, S. P., Wachs, T. D., Gardner, J. M., Lozoff, B., Wasserman, G. A., Pollitt, E., et al. (2007). Child development: Risk factors for adverse outcomes in developing countries. *Lancet*, 369, 145-157.
- Wechsler, D. (1949). *Wechsler Intelligence Scale for Children*. New York: Psychological Corporation.
- Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children-Revised*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1991). *Wechsler Intelligence Scale for Children - Third Edition*. San Antonio, TX: Psychological Corporation.

- Wechsler, D. (1997). *Wechsler Adult Intelligence Scale, Third Edition*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (2004). *Wechsler Intelligence Scale for Children - Fourth Edition*. San Antonio, TX: Psychological Corporation.
- Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage Publications.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III*. Itasca, IL: Riverside.