View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Tilburg University Repository

TILBURG ◆ UNIVERSITY

Tilburg University

There is no place like home

Malda, M.

Publication date:
2009

Link to publication in Tilburg University Research Portal

Citation for published version (APA):
Malda, M. (2009). *There is no place like home: On the relation between culture and children's cognition*. Ridderprint.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 12. May. 2021

There is no Place Like Home: On the Relation Between Culture and Children's Cognition

Maike Malda

© Maike Malda, 2009
Cover design: Judith Sebregts
Printed by Ridderprint Offsetdrukkerij B.V., Ridderkerk
ISBN/EAN: 978-90-5335-214-4

There is no Place Like Home: On the Relation Between Culture and Children's Cognition

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit van Tilburg,
op gezag van de rector magnificus,
prof. dr. Ph. Eijlander,

in het openbaar te verdedigen ten overstaan van een
door het college voor promoties aangewezen commissie

in de aula van de Universiteit
op vrijdag 27 november 2009 om 10.15 uur

door

Maike Malda
geboren op 20 juni 1981 te Tholen

Promotor: Prof. dr. A. J. R. van de Vijver

Promotiecommissie: Prof. dr. M. Q. Temane

Prof. dr. H. Keller

Prof. dr. H. van der Flier

Prof. dr. P. H. Vedder

Prof. dr. Y. H. Poortinga

Contents

Chapter 1

The Relation Between Culture, Cognition, and Bias: An Introduction

Some time ago I overheard Esther and Saskia discussing and comparing their scores on a fourth grade spelling test. Prathima also heard the two Dutch girls talking, but did not fully understand what was said. She moved with her family to the Netherlands last summer and is struggling to understand and speak Dutch. Even though Esther, Saskia, and Prathima are of the same age and attend the same grade, one immediately senses that it would be unfair to directly compare Esther's or Saskia's scores with Prathima's scores on a Dutch spelling test. Prathima would most likely perform poorer because she does not master the language, but does that mean that she is worse at spelling? Not necessarily; she might be a skilled speller in her mother tongue. What about comparing scores on a mathematics test? Every school going child learns to make calculations, but Prathima is not necessarily familiar with the same type of mathematics exercises and tests as the other girls. This could adversely affect her performance. When directly comparing school performance can already be problematic for children in the same classroom, how can we then compare a Dutch child's test performance to that of a child living in India? Bias is the name for the cause of such comparability problems and plays a key role in this dissertation.

The broad underlying question I address in this dissertation is how culture and children's cognition are related. Bias-related issues in assessing cognition and in assessing characteristics of a child's home environment play an important role in studying this relation. Taking into account bias, I shed light on the link between the concepts of culture and cognition from three angles. The first focuses on detecting and reducing bias in cognitive tests to obtain a better understanding of the relation between these concepts. Is a cognitive test suitable outside of the context in which it was originally developed? How do we know if a test is biased?

Knowing that a test is biased, how can this test be made appropriate for the target group? For the second angle, I extend the examination of cultural influences on the appropriateness of cognitive tests to cultural influences on the relation between a child's home environment and cognition. How does the home environment of a child relate to a child's cognitive performance and to what extent are theoretical models on this relation (usually developed in Western countries) applicable in a non-Western context? Whereas the first two angles focus on minimizing bias, the third and last angle focuses on experimentally manipulating bias. Knowing how to detect and reduce bias also provides information on how to introduce and manipulate bias. How does content familiarity influence cognitive test performance? Why are cross-cultural differences larger on some cognitive tests than on others? To answer these specific questions, I take the issue of bias to another level by experimentally manipulating its presence for various cognitive tests and administering the resulting test versions to various (cultural) groups.

This dissertation adds to the existing literature in three ways. First and foremost, it integrates the concepts of culture, cognition, and bias. Bias is not merely treated as a measurement problem that should be avoided; it is anticipated on, examined, and manipulated by including it in the study design. By taking bias into account rather than avoiding it, a better understanding of the relation between culture and cognition can be obtained. Second, this dissertation systematically addresses this relation from various angles. Third, it addresses the link between culture and cognition in two contexts that provide excellent conditions to do so: India and South Africa. Both countries are multicultural and do not belong to the Western countries in which most cognitive tests are developed. Many tests that are used in India and South Africa have been imported from these Western countries. The larger the cross-cultural differences between the original and target context, the more potential sources of bias, providing good conditions for a critical test of the (cross-cultural) applicability of cognitive tests and of models relating a child's direct living environment to cognitive performance.

The practical relevance of this dissertation lies in three aspects. First, it contributes to an understanding of the nature of cross-cultural score differences. Second, it provides guidelines for adequate, culture-informed test development. Appropriate instruments to assess child outcomes are needed to implement meaningful interventions that foster child development. In addition, we need to know which factors in the (day to day) environment of the child positively or negatively relate to these outcomes to identify targets for such interventions. The third practical contribution of this dissertation is the examination of (the cross-cultural validity of) the relations between certain variables in the home environment and a child's cognitive performance.

Three terms are repeatedly used throughout the text: culture, cognition, and bias. Let us first see what they actually mean, before turning to the content of the remaining chapters.

Culture

What is culture?

Numerous definitions of "culture" have been reported, illustrating the difficulty of capturing the term's meaning (Berry, Poortinga, Segall, & Dasen, 2002; Segall, Dasen, Berry, & Poortinga, 1999). Tylor (1871) was the first to provide a definition from an anthropological perspective: "that complex whole which includes knowledge, belief, art, morals, laws, customs, and any other capabilities and habits acquired by man as a member of society" (p. 1). Based on their review of definitions of culture, Kroebner and Kluckhohn (1963) arrived at the following definition:

> *Culture consists of patterns, explicit and implicit, of and for behavior acquired and transmitted by symbols, constituting the distinctive achievement of human groups, including their embodiments of artifacts; the essential core of culture consists of traditional (i.e., historically derived and selected) ideas and especially their attached values; culture systems*

> *may, on the one hand, be considered as products of actions, on the other as conditioning elements of further action.* (p. 357)

Culture is in every person and can be seen at different levels; from high (implicit and abstract) to low (explicit and concrete). On the highest level, culture reflects ideologies, norms, and values. This high level is operationalized in lower levels such as the child's direct living environment (e.g., education, language, daily activities, parental behaviors, and toys). This dissertation mainly focuses on the lower levels of culture that are more directly related to a child's cognitive performance.

Distinguishing between culture, society, and country

It might seem hard to distinguish the term culture from the term society; yet, the two are not the same (Berry et al., 2002). "Society" can be defined as "people who interact in a defined space and share culture" (Macionis & Plummer, 1998, p. 66). Society then concerns a group of interacting people, whereas culture concerns their way of life. Neither a society nor a culture coincides with the borders of a country. People speak of "Western society", which is not limited to one particular country. Also, there is no such thing as "*the* Indian culture"; all India's languages, religions, beliefs, and other explicit or implicit references to a way of life, cannot be subsumed under one common denominator that carries a significant meaning. The boundaries of a culture or society seem to depend on the level of concreteness by which these terms are examined; the more specific the aspects of interest, the stricter the boundaries. This implies that what is referred to as culture can differ across studies, making it important to be aware of a study's frame of reference.

Absolutism, universalism, and relativism

An important distinction in light of this dissertation is the one between cultural absolutism, universalism, and relativism (Berry et al., 2002). Absolutism is based on the assumption that all psychological processes and the way they are expressed are universal. According to this perspective, scores or characteristics of people across the globe can be directly compared without worrying about confounding factors; culture does not play a role. Universalism is based on the assumption that underlying psychological processes are universal, but the ways in which they are expressed are context-dependent. From a relativistic viewpoint, cultural characteristics are described without references to other cultures; they are observed from the viewpoint of one particular culture. In this dissertation, I adopt a universalistic approach. Cognitive abilities are assumed to be universal; the cultural context determines the importance attached to them and their manifestations.

Cognition

What is cognition?

Cognition can be defined as "the mental process of knowing, including aspects such as awareness, perception, reasoning, and judgment" ("The free dictionary", n.d.). Other examples of cognitive activities are attending, learning, thinking, and remembering. A cognitive task is "any task in which correct or appropriate processing of mental information is critical to successful performance" (Carroll, 1993, p. 10). A cognitive ability is then any ability "that concerns some class of cognitive tasks" (Carroll, 1993, p. 10). Memory span, word fluency, reading comprehension, and visualization are all examples of cognitive abilities.

Theories on cognitive abilities

Various theories have been described on the structure of cognitive abilities (see Flanagan & Harrison, 2005). The studies in this dissertation employ the Cattell-Horn-Carroll (CHC) model of cognitive abilities, which integrates two models. The first is the Cattell-Horn Gf-Gc theory, distinguishing between fluid and crystallized cognitive abilities. Fluid abilities reflect the "use of deliberate and controlled mental operations to solve novel, 'on-the-spot' problems (i.e., tasks that cannot be performed automatically)" (McGrew, 2005, p. 151). Crystallized abilities are "typically described as a person's wealth (breadth and depth) of acquired knowledge of the language, information and concepts of a specific culture, and/or the application of this knowledge" (McGrew, 2005, p. 151). The second model is Carroll's three-stratum theory (Carroll, 1993; McGrew, 2005), which resulted from factor analyzing over 460 cognitive data sets and describes three levels (three strata) of cognitive abilities. Abilities on the first level (Stratum I) are specific cognitive abilities, such as visual memory and spatial scanning. Factor analyzing all Stratum I abilities results in eight broad ability factors (Stratum II); examples are general memory and learning, and broad visual perception. The eight broad ability factors load on one general cognitive ability factor, which reflects the highest level (Stratum III).

Cognition versus intelligence

The reader might wonder why I am not using the term "intelligence" rather than "cognition" to describe one of the key topics of this dissertation. At first glance it seems as if these terms are interchangeable and "intelligence" seems to be a more accessible term, requiring less explanation. This is not the case, though. I intentionally do not use "intelligence" here. Ever since the term was first described in a psychology text by Herbert Spencer in the second half of the 19th century, it has been the subject of much debate (Wasserman & Tulsky, 2005). One of the most cited definitions of "intelligence" comes from David Wechsler (Wechsler, 1939):

> *Intelligence is the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment. It is global because it characterizes the individual's behavior as a whole; it is an aggregate because it is composed of elements or abilities which, though not entirely independent, are qualitatively differentiable.* (p. 3)

According to this definition (and many others), intelligence is not confined to one area of an individual's thoughts or behaviors. Western views of intelligence focus more on a person's cognitive skills (Van de Vijver & Poortinga, 2005), whereas studies in Zambia (Serpell, 1993), Japan (Azuma & Kashiwagi, 1987), and Kenya (Grigorenko et al., 2001) show that descriptions of an intelligent person go beyond the school-oriented domain that is commonly associated with intelligence in the U.S. and Europe. Social aspects are usually more relevant in everyday definitions in non-Western countries (Srivastava & Misra, 2001). The cross-cultural differences in emphasis on particular aspects of intelligence make the term ambiguous. The current dissertation specifically focuses on cognition. Unlike "intelligence", which implies a subjective evaluation of purposefulness, rationality or effectiveness of dealing with the environment, "cognition" refers to more basic processes.

Bias

What is bias?

One of the goals in cross-cultural research is to use measures and methods that are valid in the context in which they are applied. In this dissertation I consider bias as a generic term for all kinds of factors that threaten the validity of comparisons between cultural groups (Van de Vijver & Hambleton, 1996). Test 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 test bias: *construct bias*, *method bias*, and *item bias* (Van de Vijver & Poortinga, 2005; Van de Vijver & Tanzer, 2004).

Construct bias, method bias, and item bias

An instrument that shows construct bias in a cross-cultural comparison does not measure the same psychological concept across cultures. The earlier described problems with the definition of intelligence form a good example. Method bias refers to sources of bias that arise from methodological aspects of a study. There are different types of method bias, two of which are relevant for the studies addressed here. The first is instrument bias, which occurs when (cultural) groups show differential familiarity with stimulus materials (e.g., geometric shapes) or response procedures and styles (e.g., multiple choice response format). The second relevant type of method bias is administration bias. Examples are differences in environmental administration conditions (e.g., lack of comparability of testing rooms), differences in expertise of test administrators, and communication problems between administrator and participant. Item bias (differential item functioning) refers to item-specific problems in cross-cultural comparisons, such as item ambiguity due to poor item translations. Culture-specific elements can also be a source of item bias (e.g., an item about a vacuum cleaner is biased against cultures in which this appliance is uncommon).

How can we reduce bias? Adoption, assembly, or adaptation

Three terms have been proposed to describe the transformations that may be needed to transfer an instrument to another culture: *adoption (*or *application), assembly, and adaptation* (Van de Vijver, 2003). Adoption refers to a close translation into the target language. Assembly involves the construction of an entirely new instrument (Harkness, Van de Vijver, & Johnson, 2003). Adaptation has features of both adoption and assembly; it combines a close translation of the parts of the instrument that are assumed to be adequate in the target culture, such as test instructions, with a change of parts for which linguistic, cultural, or psychometric reasons make a close translation inadequate (Hambleton & De Jong, 2003; Harkness, Mohler, & Van de Vijver, 2003).

The choice of procedure depends on the translatability (Van de Vijver & Poortinga, 2005) and the expected suitability of the original instrument in the target culture (Hambleton & Patsula, 1998). Adoption of an instrument can be used if the purpose of a study is to compare scores across cultures directly (Van de Vijver, 2003); however, adoption may ignore relevant features of the target culture. For instance, a Western test of intelligence can overlook important social aspects that are part of the definition of intelligence in a non-Western context. The cultural appropriateness of the instrument in the target culture may be enhanced by using assemblies and adaptations (Hambleton & Patsula, 1999), although these two techniques offer less scope for cross-cultural comparability of scores. Assembly is applied when the original and target culture differ to such an extent that hardly any aspect of the original instrument can be retained, or when the study concerns a new research topic for which no suitable instrument is available yet. In case of an assembly, there is no identical instrument available in another culture which entirely rules out the possibility to compare scores. In case of an adaptation, these problems can be remedied to some extent by using sophisticated statistical (a posteriori) procedures that enable comparisons even when instruments are not entirely identical, such as item response theory and structural equation modeling (Van de Vijver & Leung, 1997). Adaptation is the main method of transfer in this dissertation. One of the core questions in each of the following chapters is how to make an existing Western-based assessment instrument appropriate for use in an entirely different cultural context, which requires much flexibility in changing items, while retaining the basic structure.

Integrating culture, cognition and bias

In this dissertation I show that bias is inevitably part of studies on the relation between culture and cognition. Figure 1.1 shows a simplified model relating the concepts of culture, cognition, and bias. The figure shows that culture (in the broadest sense of the word) determines the extent to which assessment instruments are biased, and this bias in turn affects the measurement of a lower level of culture, namely, the direct environment of the child (home, school, peers). Bias also affects the measurement of the relation between a child's environment and cognition, and of cognition itself. Ideally, a child's actual cognitive abilities and cognitive abilities as reflected by test scores are interchangeable. Unfortunately, when there is bias in the assessment instruments, the two are not necessarily identical. In the absence of bias, the influence of culture on cognition is confined to the salience of certain factors in the environment of the child, of certain environment-cognition relations and of certain cognitive abilities. Cognitive abilities and many environmental factors are assumed to be universal; the cultural context determines the importance attached to them and hence their salience.

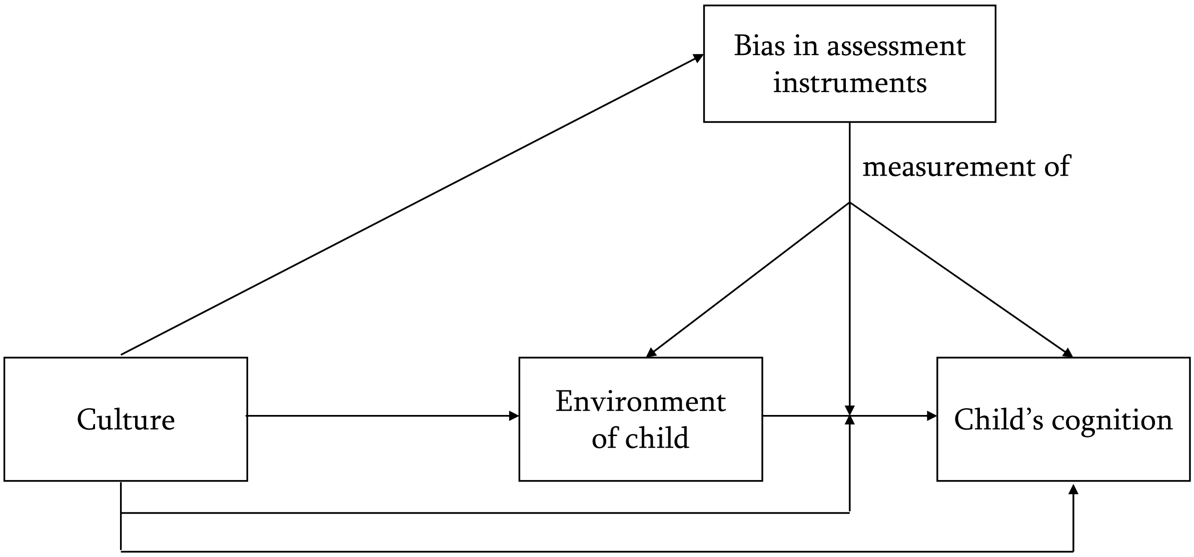


Figure 1.1
A model of culture, cognition, and bias

This dissertation

Who are studied?

This dissertation focuses on children in middle childhood, between 6 and 10 years old, from primary school grades 2 to 5. Primary schooling forms the basis for (decisions on) further education, which makes this phase very important in a child's educational career. This importance results in a high need for the awareness and reduction of bias in assessment instruments, so as to obtain a valid estimation of a child's cognitive abilities. Only children with at least one year of formal education are included in the studies to ensure some skill development needed in cognitive testing procedures, such as language skills, understanding of instructions, and familiarity with performing on tasks. I confine the sample to children in middle childhood because the studies do not have the purpose of identifying developmental trends across age groups.

Overview of chapters

Figure 1.1 shows that bias can affect the measurement of cognition. A cross-cultural comparison of scores on a cognitive test that is biased against at least one of the tested groups is invalid. How do we know if a test is biased? What is the next step if we know that a test is biased? How can this test be made appropriate for the target group? I aim to answer these questions in Chapter 2.

There are two types of procedures to detect and reduce bias, namely a priori and a posteriori procedures. The former involve a qualitative, judgmental approach, such as piloting test items to investigate their cultural appropriateness and adapting items accordingly, before data collection starts. The latter are statistical procedures and are applied after data collection. Chapter 2 focuses on a priori procedures in adapting 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. Many guidelines exist on how adaptations should be made (American Educational Research Association, American Psychological Association, & National Council on Measurement of Education, 1999; Hambleton, 2001); however, a combination of these theoretical guidelines and a detailed illustration of their application has hardly been described. Chapter 2 proposes and applies a systematic, qualitative approach to adapt cognitive tests. The adapted KABC-II results from a pilot study with 57 children.

After the qualitative (non-empirical) procedures to increase the cultural suitability of a cognitive instrument as described in Chapter 2, the next question is how the appropriateness of the adaptation can be statistically checked. Is an adaptation of a Western cognitive instrument for a non-Westernized resource-limited setting reliable and valid? Chapter 3 focuses on this question. Quantitative criteria to assess the quality of an adaptation include the instrument's validity and reliability. Chapter 3 tests whether the adapted version of the KABC-II meets three validity criteria; first, the theoretical model underlying the original instrument (the CHC-model) should be well represented in the data; second, relations of test scores with demographic variables such as children's age and sex should be according to expectations; third, relations of test scores with presumably related psychological constructs (such as scholastic achievement) should be according to expectations.

Reliabilities of all adapted subtests are addressed as well. Data are collected for 598 children 6- to 10-year-old Kannada-speaking children of low socioeconomic status in Bangalore, India.

Chapters 2 and 3 describe a qualitative approach and a following quantitative approach to ensure the cultural appropriateness of a cognitive test adaptation. I extend the examination of cultural influences on the appropriateness of cognitive tests to cultural influences on the relations between a child's home environment and cognition in Chapter 4. As shown in Figure 1.1, culture can influence the salience of certain environment-cognition relations and culture can induce bias in measuring them. To what extent are Western models on these relations applicable in a non-Western context? The structure of these links is expected to hold across cultures; yet, both their significance and strength may vary. Chapter 4 examines the suitability of the combined Western investment model and the family stress model (Guo & Harris, 2000; Yeung, Linver, & Brooks-Gunn, 2002) for studying relations between home factors and children's cognitive outcomes in families of low socioeconomic status in Bangalore. The investment model holds that if more financial resources are invested in children, children will have access to more materials and activities that can enhance their cognitive performance. The family process or family stress model (Conger et al., 2002) relates low income to material hardship and resultant parental stress, which can affect child development through mediating variables, such as parental behaviors. Both models can be viewed from a proximal-distal perspective (Bronfenbrenner, 1979), reflecting the relative proximity of environmental variables to child outcomes. Socioeconomic status can be seen as a variable with a distal, more general influence on these outcomes, whereas parental behaviors are more proximal, with a focused influence. The primary caregivers of 532 children that underwent the KABC-II test administration (see Chapter 3) are interviewed. Questionnaires cover socioeconomic status, general mental health, perceived social support, family conflict, and parental behaviors. The relations between all variables are tested in a path model.

After focusing on the cross-cultural validity of a Western cognitive test (Chapters 2 and 3) and of Western models relating home environment to cognitive performance (Chapter 4), Chapter 5 takes the issue of bias to another level by experimentally manipulating it. Why are larger cross-cultural differences found on some cognitive tests (fluid reasoning tests) than on others (short-term memory and attention tests)? Spearman's Hypothesis (SH) relates these cross-cultural differences to the cognitive complexity of tests; differences are larger for tests with a higher cognitive complexity (Jensen, 1985, 1998). SH assumes that there are cross-cultural differences in the underlying general cognitive ability on which tests with higher cognitively complexity more strongly rely. I believe, however, that it is cultural complexity rather than cognitive complexity that explains these differences (Helms-Lorenz, Van de Vijver, & Poortinga, 2003) and expect fluid reasoning tests to be more sensitive to cultural information than other (more basic cognitive) tests. Cultural complexity is conceptualized in the present study as the extent to which test content (i.e., words, drawings) is more familiar to one of the compared groups. In Chapter 5, the content familiarity of five cognitive tests (for short-term memory, attention, working memory, and figural and verbal fluid reasoning) is manipulated to examine its differential effect on test performance for three groups of South African children: 161 Afrikaans, 181 urban Tswana, and 159 rural Tswana children in the third and fourth grade of primary school.

Chapters 2 to 5 are based on published and submitted articles that can be read independently of one another. Chapters 2, 3, and 4 are part of one project, which implies some overlap in content. Chapter 6 summarizes the reported findings and discusses these in light of implications for culturally appropriate test development.

Chapter 2

Adapting a Cognitive Test for a Different Culture: An Illustration of Qualitative Procedures[1]

...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 an absence 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 cognitive abilities 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, & Puhan, 1997). Since cognitive test scores are known to predict school performance of children (also in non-Westernized countries), it is important for

[1] This chapter is based on Malda et al. (2008)

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 in 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 (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).

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 cognitive 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 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.

Adoption, assembly, and adaptation

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), assembly,* and *adaptation* (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. First, the instrument shows generalizability to other, widely used cognitive batteries regarding instruction, item, and response formats. Second, 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 pretesting) 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 choice 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. 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).

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 and *Pattern Construction* from the Differential Ability Scales (Elliott, 1990).

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. First, 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. Second, 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). 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, one item from WISC-III was used, 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 breaking 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 2.1).

Figure 2.1
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 one-syllable 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 one-syllable 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 to the child out loud at a rate of one word per second 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 two times 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. First, we systematically employed iterations of translating, piloting (i.e., cognitive pretesting), and modifying items. Second, 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) 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 on our KABC-II adaptation and on cognitive test adaptations in general. First, 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.

Second, 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? First, 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. Second, 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. Third, 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). Fourth, 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 cognitive 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. The questions need to be addressed of whether the adapted subtests constitute reliable and valid measures. In short, the data collection provides the litmus test of the adequacy of the adaptation (see Chapter 3). 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.

Chapter 3

Traveling With Cognitive Tests: Testing the Validity of a KABC-II Adaptation in India[2]

Many tests that originate from the U.S. or Europe have been used extensively in developing and emerging countries without practically and scientifically examining the suitability of these instruments outside of their country or culture of origin (Misra, Sahoo, & Puhan, 1997). The use of the original or closely translated instruments saves costs and time; yet, optimizing an instrument for a specific cultural context implies the need for an adaptation, in which cultural knowledge, values, and practices are taken into account (see Abubakar et al., 2007; Holding et al., 2004). A close translation may then not be sufficient. The validity of adapted instruments cannot be inferred from the original Western instruments and has to be demonstrated in the new cultural context. Adaptations of the Kaufman Assessment Battery for Children (K-ABC; A. S. Kaufman & Kaufman, 1983) have been described before (e.g., Holding et al., 2004; Moon, McLean, & Kaufman, 2003); however, its successor, the Kaufman Assessment Battery for Children, second edition (KABC-II), which differs from the first version in several ways (see A. S. Kaufman & Kaufman, 2004), has not been adapted. The KABC-II is an individually administered measure of cognitive ability that can be used for children from 3 to 18 years of age. It measures long-term storage and retrieval, short-term memory, visual processing, fluid reasoning, and crystallized ability. In a previous report we described the extensive adaptation of the KABC-II for use among 6- to 10-year-old Kannada-speaking children of low socioeconomic status in Bangalore, South India (Chapter 2). The present study statistically tested the adequacy of the resulting instrument by examining its reliability and validity.

[2] This chapter is based on Malda, Van de Vijver, Srinivasan, Transler, & Sukumar (in press)

Quality criteria

The use of Western cognitive instruments in non-Western contexts may lead to bias, which refers to factors that make direct comparisons of test constructs or scores between groups invalid (e.g., Van de Vijver & Tanzer, 2004). Bias makes culture-fair testing impossible (Verney, Granholm, Marshall, Malcarne, & Saccuzzo, 2005). Instrument adaptation has been proposed as a strategy to reduce bias, optimize the suitability of an instrument for a cultural context, and facilitate cross-cultural test transfer (e.g., Van de Vijver & Poortinga, 2005). A properly adapted cognitive instrument meets both qualitative and quantitative criteria.

The main qualitative criteria involve cultural and linguistic appropriateness of the instrument in the target context. The judgmental procedures that were applied to meet these criteria in the Indian adaptation of the KABC-II were described before (see Chapter 2). Quantitative criteria to assess the quality of an adaptation include the instrument's reliability and validity. Various criteria have been proposed for demonstrating the construct validity of an instrument (Messick, 1989). In the absence of cross-cultural comparative data, our validity test was based on three criteria that examine whether theoretical expectations are borne out. First, the underlying theoretical model should be well represented in the data; second, relations of test scores with background characteristics such as children's age and sex should be according to expectations; thirdly, relations of test scores with presumably related psychological constructs (such as scholastic achievement) should be according to expectations. The current study tested whether the adapted version of the KABC-II meets these three criteria that are described below in more detail. Other criteria for construct validity, such as a test's predictive validity, were not addressed.

Generalizability of cognitive structure. The KABC-II is based on the Cattell-Horn-Carroll (CHC) model of cognitive abilities (A. S. Kaufman & Kaufman, 2004; see also Reynolds, Keith, Goldenring-Fine, Fisher, & Low, 2007). The CHC theory integrates the Cattell-Horn Gf-Gc theory (distinguishing between various fluid and crystallized cognitive abilities) and Carroll's three-stratum theory (Carroll, 1993; McGrew, 2005). The construct validity of our adaptation is statistically supported if our Indian data confirm the original CHC structure, assuming that the model can be generalized to non-Western groups. The generalizability of the CHC model has been shown with exploratory as well as confirmatory factor analyses across age (Bickley, Keith, & Wolfle, 1995; Taub & McGrew, 2004) and sex (Reynolds, Keith, Ridley, & Patel, 2008). Furthermore, the CHC structure is found with many cognitive test batteries even when these were not originally designed to represent this structure (for an overview see McGrew, 2005). Is the CHC model, besides being generalizable across ages, sexes, and tests, also generalizable across cultures?

Models underlying cognitive test batteries have shown cross-cultural stability (Berry, Poortinga, Segall, & Dasen, 2002; Irvine, 1979; Van de Vijver, 1997). A large study was conducted with culturally adapted versions of the WISC-III (Georgas, Van de Vijver, Weiss, & Saklofske, 2003). Using exploratory factor analyses, it was found that the cross-cultural equivalence of the underlying structure was high. The K-ABC (the predecessor of the KABC-II), based on the sequential versus simultaneous processing distinction, was applied in many non-Western countries (e.g., Boivin et al., 1996; Conant et al., 1999; Jansen & Greenop, 2008; Mardell-Czudnowski, 1995). Holding et al. (2004) and Moon et al. (2003) found the underlying model to be present in adapted versions in Kenya and Korea, respectively. Not many studies on the CHC model have been conducted in non-Western contexts; still, there is no reason to doubt the universality of the structure of a well-established cognitive model. We consider the CHC model to be a good starting point for statistically evaluating the validity of our KABC-II adaptation.

Age and sex effects. Test scores are expected to increase with age in our study sample. Although there tends to be considerable overlap between male and female cognitive test score distributions (Born, Bleichrodt, & Van der Flier, 1987; Fairweather, 1976), males generally score higher on tests measuring visual abilities and mathematical reasoning, whereas females do better on verbal (memory) tasks and numerical calculation (e.g., Reynolds et al., 2008). Sex differences on the original KABC-II are small for school-age children (7 to 18 years); boys perform better on the visual processing tasks and girls on the learning and fluid reasoning tasks (A. S. Kaufman & Kaufman, 2004).

Relations with school performance: Arithmetic. Various mechanisms behind the positive relation between cognitive abilities and arithmetic skills in middle childhood have been proposed, such as phonological and/or visuospatial memory, speed of processing, number processing, and spatial and non-verbal ability (Durand, Hulme, Larkin, & Snowling, 2005). The mechanisms are likely to vary across ages and tasks. Correlations between broad ability factors of the original KABC-II and arithmetic scores of other cognitive batteries (A. S. Kaufman & Kaufman, 2004) show that for the younger age group (grades 2-5) the highest correlations were found for the fluid reasoning factor, possibly because arithmetic processes are not yet differentiated and automated, and hence, their solution still requires complex, integrated cognitive abilities. We expect similar findings in the present study.

Hypotheses

Our study adds to the literature in that it 1) evaluates the validity of an adaptation of the relatively new KABC-II in a non-Western context; and 2) examines the CHC model in this non-Western (Indian) context to accomplish this. As a prerequisite for any hypothesis testing in a research context (rather than a clinical context), the internal consistencies of the subtests should be at least .70 (Cicchetti et al., 2006). The appropriateness of our adaptation is tested using the following hypotheses:

- Theoretical structure
    1. The factor structure of the KABC-II adaptation is in line with the CHC model;
- Psychometric properties

2. The underlying cognitive structure is similar across sexes (2a) and ages (2b);
3. Test scores increase significantly with age;
4. If sex differences in scores are found, boys outperform girls on visual processing tasks, and girls outperform boys on fluid reasoning and learning tasks;
5. All broad ability factors correlate significantly with arithmetic scores;
6. Arithmetic scores show the highest correlations with the fluid reasoning factor.

Method

Participants and study context

The sample included 598 Kannada-speaking children (293 boys and 305 girls) of low socioeconomic status in Bangalore (state of Karnataka, South India). The children were between 6 and 10 years old ($M = 8.71$, $SD = 1.17$) and from grade two to five of two primary schools ($N = 370$ and $N = 228$, respectively).

The children in our sample came from families with an average monthly income of 2700 Indian Rupees (56 USD). Many adults were illiterate or had only a few years of education. Houses were crowded; most had one or two rooms, and the average number of people in a household was 5.81 (2.71 adults and 3.11 children). Children had very few toys to play with and very limited access to books. Rote learning was widely used in the schools of the study; it is a commonly applied method in Indian education (Mishra, 1997), which is well applicable with large numbers of children and with a collectivistic style of teaching because the children do not need to be addressed individually.

Instruments

Kaufman Assessment Battery for Children, Second Edition. An adapted version of eight core subtests of the KABC-II was administered; a description of the extensive adaptation procedure can be found in Chapter 2. Here, we confine ourselves to describing the abilities measured (see also Carroll, 1993; A. S. Kaufman & Kaufman, 2004; J. C. Kaufman, Kaufman, Kaufman-Singer, & Kaufman, 2005; McGrew, 2005). Two subtests were selected to assess the broad ability of fluid reasoning; Pattern Reasoning measures inductive reasoning and visualization, and our Indian version of WISC/-R/-III Picture Arrangement (which is a replacement of the KABC-II subtest Story Completion) mainly measures pattern recognition, reasoning, and planning (Wechsler, 1949, 1974, 1991). Two subtests were selected to cover the broad ability factor of short-term memory, namely Number Recall and Word Order, both measuring memory span. The subtests Rover and Triangles were assumed to reflect the broad ability of visual processing. Rover is a measure of spatial scanning, general sequential or deductive reasoning, and math achievement; Triangles measures spatial relations and visualization. For the broad ability factor of long-term storage and retrieval, Atlantis was selected which is a measure of associative memory. The second test was an auditory Verbal Learning Test, which was a replacement of the KABC-II subtest Rebus. In this study we use the recall score of the Verbal Learning Test: the number of words (out of 15) correctly recalled after a 20 minute delay. For the purposes of the larger study to which our study contributes, another subtest that is not part of the KABC-II battery was added, namely Verbal Fluency (a measure of associational fluency). This addition aimed to ensure proper coverage of long-term storage and retrieval processes. Also, two subtests reflecting the broad ability of cognitive speediness were included, namely Coding B and Number Cancellation; cognitive speediness was not covered by the core subtests of the KABC-II but is part of the CHC model. Coding B is taken from the Wechsler scales and is (mainly) a measure of attention and concentration; Number Cancellation measures perceptual speed, and more specifically scanning (McGrew, 2005).

Arithmetic test. Measures of crystallized abilities and school achievement (such as reading, spelling, and arithmetic tasks) provide an external validation criterion for our adapted test. The only available local test of school achievement that seemed sensitive enough to discriminate between children within the same school grade, was the Arithmetic Diagnostic Test for Primary School Children (Ramaa, 1994). We based our arithmetic test on this instrument and on information about the curriculum of the schools included in the current study. A test consisting of two parts was composed; 36 addition and 34 subtraction items of increasing difficulty. The sum score of these two parts was used in the analyses.

Procedure

All children and their parents gave consent for participation in the study, in accordance with Indian ethical rules. Seven test examiners were trained jointly by a Dutch psychologist (first author) and a local psychologist (fifth author), after which the study commenced. Each test examiner assessed two children every day; the administration was split up in three test sessions (two sessions of 30 minutes and one of 45 minutes). All administered the tests to more or less the same number of girls and boys and to children of all grades. The arithmetic test was administered to batches of about 60 children who had all received KABC-II testing in the same week.

Analyses

First, reliabilities were calculated by the split-half technique, Cronbach's alpha, or correlations, depending on the characteristics of the subtests. Second, structural equation modeling in Amos 6 (Arbuckle, 2005) was used to test the validity of the CHC model for the study sample (hypothesis 1), followed by multigroup analyses to test for equivalence of the model across sexes (hypothesis 2a) and ages (hypothesis 2b). Multivariate Analyses of Variance (MANOVAs) were conducted to test for the effects of age (hypothesis 3) and sex (hypothesis 4) on cognitive test scores. Last, hypotheses 5 and 6 (dealing with the arithmetic test) were addressed by correlating broad ability scores with arithmetic scores, using Pearson correlation coefficients.

Results

Reliability

For the subtests Atlantis, Rover, Number Recall, Pattern Reasoning, Word Order, Triangles, and Picture Arrangement, the internal consistency was measured by the split-half technique. Values of Cronbach's alpha could not be computed due to the discontinuation rules of these subtests. For each subtest, the sum scores of the odd and even items were correlated and the Spearman-Brown formula (Thurstone, 1931) was applied to adjust this reliability estimate for test length. Reliabilities of our adapted subtests were acceptable to very good (Pattern Reasoning: .94; Picture Arrangement: .72; Number Recall: .70; Word Order: .82; Triangles: .89; Rover: .90; Atlantis: .96) and largely in accordance with the reliabilities of the original KABC-II.

The Verbal Learning Test comprised a 15-word list that was read out loud to the child, after which immediate recall was measured. This procedure was repeated twice and after a 20 minute delay, recall was measured for the fourth time. Cronbach's alpha was calculated for the number of correctly recalled words in each of these four trials; the median alpha across the five age groups (i.e., ages 6 to 10) was high: .84 (range .75 - .86). In further analyses we only used the recall score of the Verbal Learning Test (i.e., number of correctly recalled words after a 20 minute delay).

For Verbal Fluency, Number Cancellation, and Coding B, an indication of reliability could only be obtained by item or test correlations because these tests consist of one or two items. All correlations were controlled for age. The Verbal Fluency test first required the children to call out as many animals as possible, and then as many first names as possible. The correlation between the two numbers was positive and significant ($r(598) = .31$, $p < .01$), according to our expectations. The value was not very high, presumably because most children named their classmates one by one when they generated first names, but they did not use a common strategy in generating animal names. The correlation between Number

Cancellation time and Coding B was $r(598) = -.45$ ($p < .01$), indicating that the faster the child finished the Number Cancellation task, the more correct items were obtained on Coding B.

Validity

CHC model. Structural equation modeling was used to test the validity of the CHC model. The subtests (i.e., specific abilities) were expected to cover five broad abilities, namely fluid reasoning, short-term memory, visual processing, long-term storage and retrieval, and cognitive speediness. A general cognitive ability factor (called Mental Processing Index for the KABC-II) was expected to underlie these five factors. The fit of the original CHC model (model 1) to our data was acceptable, however, the modification indices suggested two improvements: 1) linking Verbal Fluency to the cognitive speediness factor rather than the long-term storage and retrieval factor (because Verbal Fluency also involves speed), and 2) combining the visual processing subtests with the fluid reasoning subtests in one factor (because all these subtests involve aspects of reasoning). We tested the first alternative (model 2), then the second alternative (model 3), followed by a combination of the two (model 4). Fit statistics are presented in Table 3.1.

Table 3.1

Summary statistics for various structural equation models related to the CHC model

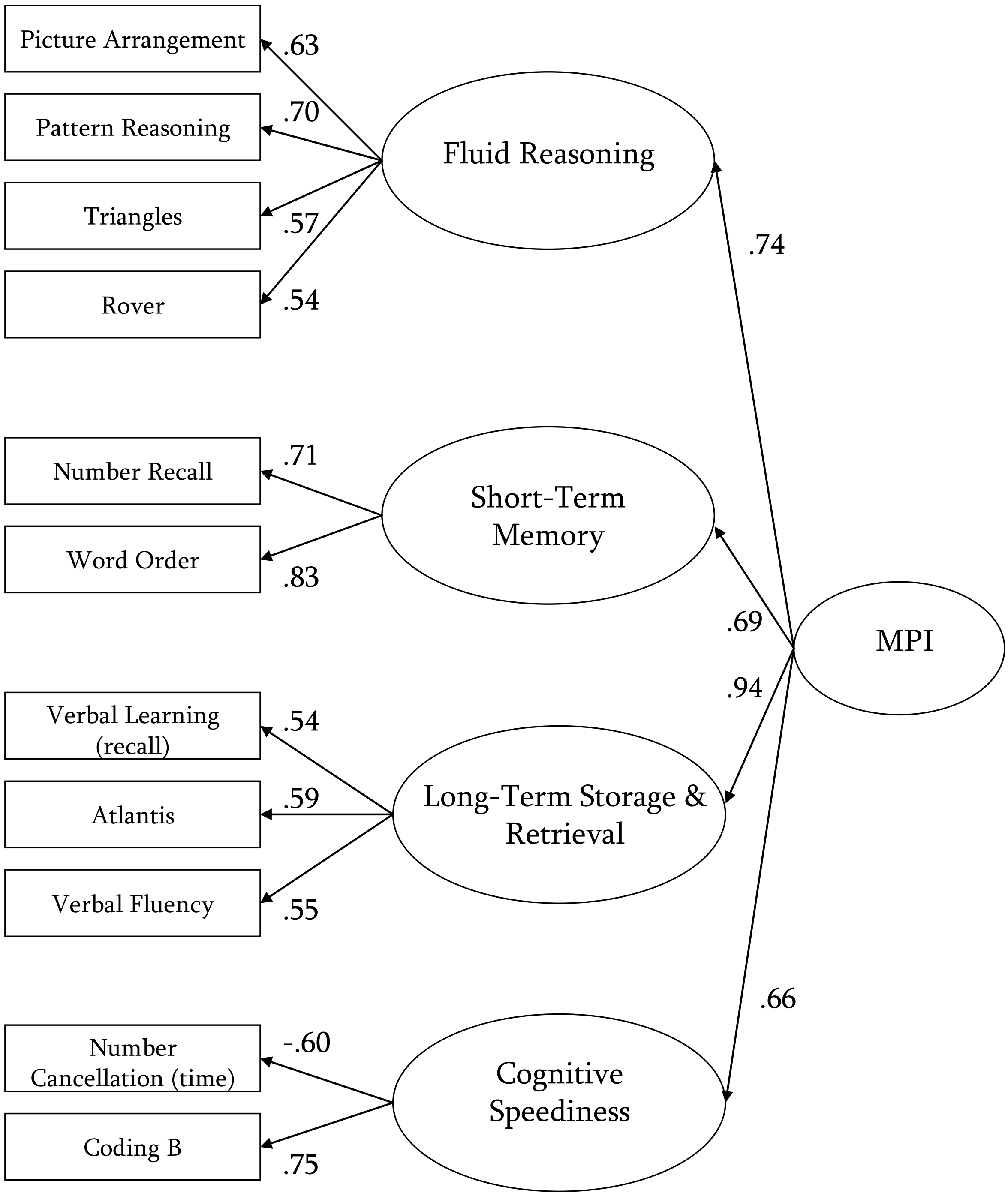
Model	χ^2	*df*	χ^2/df	*p*	CFI	RMSEA	AIC
1. CHC	122.711	39	3.146	.001	.941	.060	176.711
2. VF to CS	124.312	39	3.187	.001	.940	.061	178.312
3. VP with FR	*103.792*	*40*	*2.595*	*.001*	*.955*	*.052*	*155.792*
4. Vf to CS and VP with FR	111.406	40	2.785	.001	.950	.055	163.406

Note. Preferred model is printed in italics. CHC = Cattell-Horn-Carroll, Vf = Verbal Fluency, CS = Cognitive Speediness, VP = Visual Processing, FR = Fluid Reasoning, CFI = Comparative Fit Index, RMSEA = Root Mean Square Error of Approximation, AIC = Akaike's Information Criterion.

Differences between the fit of the models were small, however, Table 3.1 suggests that model 3 showed a slightly better fit than the other models (particularly given its relatively low AIC value). Model 3 is displayed in Figure 3.1.

There is considerable overlap between the structures of the CHC model and our model; most importantly, the hierarchy of cognitive abilities is supported. Most expected factors were found: a short-term memory factor, a long-term storage and retrieval factor, and a cognitive speediness factor. Our final model differed from the CHC model in two ways. First, Triangles and Rover, hypothesized to represent a separate visual processing factor, loaded on the same factor as Pattern Reasoning and Picture Arrangement. Second, the patterning of the loadings of the broad abilities on general cognitive functioning is different from the CHC model. This model would predict that fluid reasoning has the highest loading on the general cognitive factor, followed by short-term memory, long-term storage and retrieval, and cognitive speediness. However, the long-term storage and retrieval factor showed the highest loading in our study, followed by fluid reasoning, short-term memory, and cognitive speediness. We estimated the confidence intervals of the factor loadings using a bootstrapping procedure and found that the confidence interval of the long-term storage and retrieval loading did not overlap with the confidence intervals of the other loadings. This indicated that only the long-term storage and retrieval loading differed significantly from the others ($p < .05$).

Sex and age. Multigroup analysis was applied to test for equivalence of our final model across sexes and ages. Sex and age could not be combined in one single multigroup analysis because the children were not equally distributed across all possible combinations of sex and age; therefore, we tested multigroup invariance of the CHC model separately for these two variables. A good fit was found for a model in which all parameters were constrained to be equal for boys and girls ($\chi^2[106, N = 598] = 213.64$, $p < .01$, $\chi^2/df = 2.02$, GFI = .94, AGFI = .93, TLI = .95, CFI = .95, RMSEA = .04).



Note. Standardized coefficients. The model is standardized for age. MPI = Mental Processing Index.

Figure 3.1
Structural equation model of all cognitive data

The model testing the invariance across age groups (age groups 6 and 7 are taken together due to the relatively small number of children with age 6) showed a good fit when all parameters were identical, except for equality of measurement residuals (χ^2[205, *N* = 598] = 298.58, $p < .01$, χ^2/df = 1.46, GFI = .92, AGFI = .89, TLI = .93, CFI = .94, RMSEA = .03). Widaman and Reise (1997) argue that the latter constraint is rather irrelevant.

A MANOVA was computed with sex (Wilks' Lambda = .88, *F*[11, 578] = 7.23, $p <$ .01, partial η^2 = .12) and age group (Wilks' Lambda = .57, *F*[44, 2213.24] = 8.05, $p <$ .01, partial η^2 = .13) as independent variables and the sum scores for all cognitive tests as dependent variables. There was no significant interaction between sex and age group (Wilks' Lambda = .92, *F*[44, 2213.24] = 1.10, $p = .31$, partial η^2 = .02). Girls scored significantly higher than boys on Verbal Learning recall (*F*[1, 588] = 18.82, $p < .01$, $d = .42$), Number Cancellation time (*F*[1, 588] = 36.04, $p < .01$, d = -.47), Coding (*F*[1, 588] = 13.90, $p < .01$, $d = .38$), and Verbal Fluency (*F*[1, 588] = 8.34, $p < .01$, $d = .30$). Boys outperformed girls on Rover (*F*[1, 588] = 5.31, $p < .05$, d = -.18). Scores on all subtests increased significantly with age ($p < .01$). Age explained more variance on the subtests measuring fluid reasoning and cognitive speediness as compared to the subtests measuring short-term memory and retrieval ability. For example, for Rover, partial η^2 was .17 (*F*(4, 588) = 30.61, $p <$ .01) whereas for Number Recall, the value was .04, *F*(4, 588) = 6.05, $p < .01$. These findings indicate that children's reasoning and speed abilities are more age dependent than the other abilities.

Arithmetic test. The reliability of the arithmetic test was very high; the median value of Cronbach's alpha for the five age groups was .95 (range .93 - .96). The sum scores on the arithmetic test correlated significantly with the broad ability factors (with *r*(598) ranging from .37 to .49, $p < .01$) and the Mental Processing Index (*r*(598) = .61, $p < .01$). The correlation was significantly lower (tested using procedures described by Dunn & Clark, 1969) for cognitive speediness than for all other factors (the difference between the cognitive speediness and short-term memory factor was bordering on significance). Arithmetic scores did not show the expected stronger correlation with the fluid reasoning factor than with the other factors.

Discussion

The KABC-II was extensively adapted for 6 to 10-year-old Kannada-speaking children of low socioeconomic status in Bangalore, South India (Chapter 2). The current study statistically evaluated the adequacy of our adaptation. Most hypotheses were confirmed. The adapted subtests showed high reliabilities; the cognitive CHC model underlying the original KABC-II was largely replicated (hypothesis 1); the CHC model was valid across sexes (2a) and age groups (2b); cognitive test scores increased with age (3); the small sex differences in some of the subtest scores were in line with expectations (4); the arithmetic test correlated significantly with all broad ability factors (5); the arithmetic sum score showed similar correlations with all but one of the broad ability factors (not confirming hypothesis 6, which predicted a higher correlation with fluid reasoning). We can conclude that our adaptation of the KABC-II is a valid cognitive measure for the target sample. This is an important finding in light of the prevalence of bias in applications of Western cognitive instruments in a non-Western context.

Our final model differed from the CHC model in two ways. First, Triangles and Rover, hypothesized to represent a separate visual processing factor, loaded on the same factor as Pattern Reasoning and Picture Arrangement. These four subtests were among the subtests with the highest intercorrelations (varying from .33 to .48, $p < .01$). This finding most probably relates to a combination of the nature of the tasks (all are figural) and their complexity; fluid abilities usually depend on an integration of many distinct cognitive abilities whereas visual processing abilities require fewer cognitive resources. However, since both the stimulus mode and the response mode of our visual processing tasks were very unfamiliar to the Indian children (despite the test adaptations that were performed to increase familiarity and suitability), it stands to reason that Triangles and Rover reflect fluid reasoning abilities rather than merely visual processing. Training children on various cognitive tasks could reduce the cognitive complexity and cause a better differentiation between tests (and factors) and a smaller reliance on a more general cognitive ability.

Second, the patterning of the loadings of the broad abilities on general cognitive functioning is different from the CHC model. Only the loading of the long-term storage and retrieval factor differed significantly from the others. The limited differentiation between the loadings may, again, be a consequence of the task unfamiliarity. The high loading of the long-term storage and retrieval factor may be caused by the diversity of the subtests that belong to this factor. The factor consists of tests that tap a wide range of abilities, in addition to their common factor. The Verbal Learning Test is a test of short- and long-term memory and learning, whereas Verbal Fluency measures free recall, and Atlantis adds a strong visual component (besides its memory and learning aspects). Because together these tests measure a broad range of abilities, it is not surprising that the loading of the long-term storage and retrieval factor on the Mental Processing Index (MPI, reflecting general cognitive functioning) is so high. The CHC model might have been replicated more closely, if we had included all (instead of merely the core) subtests of the KABC-II, providing a better coverage of the broad ability factors. Adding subtests would however have led to prohibitively long test administration times for our study sample.

Familiarity could also play a role in explaining the lack of differentiation in correlations between the broad ability factors and the arithmetic score. In the introduction we suggested that arithmetic processes might not yet be differentiated and automated for young children, and hence, their solution still requires complex, integrated cognitive abilities. This suggestion is in line with our expectation of a higher correlation between the arithmetic score and the (cognitively complex) fluid reasoning factor. Our findings, however, show that the arithmetic score correlates similarly to all but one of the broad ability factors (cognitive speediness), which might indicate that the arithmetic test measures an even more cognitively complex ability in our sample than anticipated. The high correlation with the MPI (.61) confirms the relevance of general cognitive processes in arithmetic performance, possibly due to the unfamiliarity of the sample with such tests and test situations.

The task unfamiliarity that we observed in our sample shows the profound influence of both home environment and educational characteristics on cognitive test performance. These Indian children of low socioeconomic status are provided with suboptimal stimulation (few play materials) at home and the educational system is mainly focused on collective rote learning, which explains the child's lack of experience with individual test situations, and with materials such as puzzles and (geometrical) figures. This implies that issues with testing in non-Western contexts could be related to differences in socioeconomic status, in addition to cultural differences. It is possible that a sample of children from the same geographic area and same language but with high socioeconomic status would have shown a closer match with the original CHC structure compared to our present sample.

The combination of the evidence obtained in the qualitative adaptation process of the KABC-II (Chapter 2) and the quantitative process discussed here, supports the suitability and validity of our adaptation for Kannada-speaking children of low socioeconomic status in India; the current study offers further evidence for the generalizability of the CHC model in developing, non-Westernized countries. Both the qualitative and quantitative parts are prerequisites for ensuring an instrument's adequacy. Many studies omit a detailed test adaptation, which could lead to the use of culturally inappropriate stimuli. However, the current study shows that after an extensive qualitative adaptation process, quantitative analyses are needed to demonstrate its success. Cognitive data can only be interpreted validly when the tests meet both judgmental (qualitative) and statistical (quantitative) adaptation criteria.

Chapter 4

Stimulation in the Home Positively Relates to Children's Cognitive Performance in Bangalore, India[3]

Numerous studies have addressed the relation between poverty and children's developmental outcomes (see Bradley & Corwyn, 2002; Brooks-Gunn & Duncan, 1997); yet, studies of the pathways underlying this relation were mainly conducted in the United States (Bradley & Corwyn, 2002; Davis-Kean, 2005; Guo & Harris, 2000; Lugo-Gil & Tamis-LeMonda, 2008; Yeung, Linver, & Brooks-Gunn, 2002). To what extent are Western models on the structure and strength of relations between variables in the home environment and developmental outcome variables generalizable across countries or cultural groups? Relations between these variables can differ across groups for various reasons (Bradley et al., 1989; McLoyd, 1998), such as sample characteristics (e.g., different socioeconomic circumstances), cultural aspects (e.g., different norms or parenting practices), and measurement features (e.g., cultural bias in the assessment instruments). Cross-cultural differences in (the strength of) these relations could point to differences in factors that are critical for developmental outcomes. Specific targets for interventions fostering child development might therefore be dependent on the cultural context. The present study tests the applicability of Western findings on environment-cognition relations in a non-Western context. This is done by examining which variables in the home environment are associated with the cognitive performance of children of low socioeconomic status (SES) in Bangalore (South India) as well as possible pathways underlying this relation.

Conceptual framework of the present study

The study starts from an integration of two Western models that have been proposed to examine relations between SES (indicators) and developmental

[3] This chapter is based on Malda, Van de Vijver, Srinivasan, Transler, & Sukumar (2009)

outcomes of children: the investment model and the family stress model (Guo & Harris, 2000; Yeung et al., 2002). The investment model holds that if more financial resources are used to invest in children, children will have access to more materials and activities that can enhance their cognitive performance. The family process or family stress model (Conger et al., 2002) can be seen as an extension of the investment model. The model relates low income to material hardship and resultant parental stress, which can affect child development through the family climate and parental behaviors. In short, the investment model focuses on a family's material resources, whereas the family stress model focuses more on psychological resources; both types of resources can affect a child's cognitive outcomes. The variables included in these models vary over studies. We combine the two models in the current study by including a core selection of variables that reflect both material and psychological resources, namely SES, asset indicators (i.e., housing conditions and goods), psychological functioning of the primary caregiver, social support, family conflict, and parental behaviors. Figure 4.1 shows a model of the possible relations between these variables.

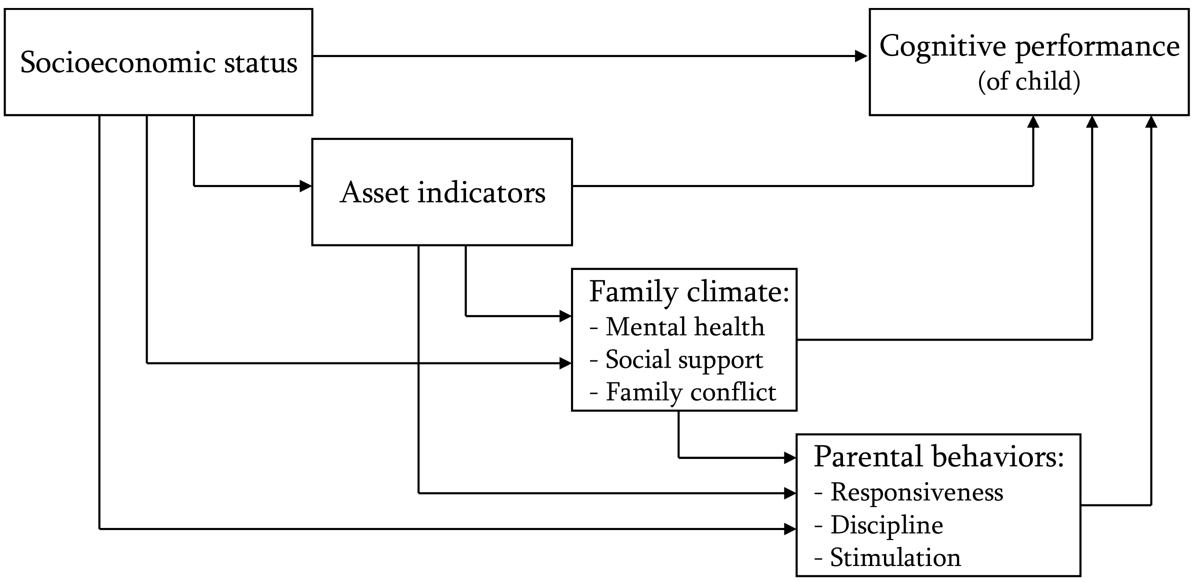


Figure 4.1
The hypothesized model relating home environment to cognitive performance

In line with other developmental models (e.g., Bronfenbrenner, 1979), we distinguish variables that are proximal from variables that are distal to a child's cognitive outcome. The variables displayed in Figure 4.1 can be viewed as varying on this proximal-distal dimension. According to this perspective, developmental outcomes are associated with a hierarchically organized network of variables that range from SES as most distal, with a general influence, to parental behaviors as most proximal, with a focused influence. Going from left to right in Figure 4.1, the variables become more proximal to the child's cognitive performance. Crucial in the model is the mediating role of more proximal variables in the association between more distal variables and child outcome. The effect of the distal variable SES on outcome is "unpackaged" (Whiting, 1976) into effects mediated by more proximal variables.

The next section describes the relations between these variables in more detail and, where necessary, explains their relevance for the Indian context. The empirical studies were performed with Western samples unless specified otherwise.

Relevant distal and proximal variables in an Indian context

SES is usually seen as a variable with a distal influence on developmental outcomes and in many studies represents variables such as parental occupation, education, and income (McLoyd, 1998). Housing conditions and household goods (i.e., asset indicators) are used as more concrete indicators of a family's SES (Filmer & Pritchett, 2001). More indirect (Gershoff, Aber, Raver, & Lennon, 2007; Guo & Harris, 2000; Krishnakumar & Black, 2002; Yeung et al., 2002) than direct (Duncan, Brooks-Gunn, & Klebanov, 1994; Krishnakumar & Black, 2002) relations have been found between (indicators of) SES and cognitive outcomes. Direct links were reported between socioeconomic variables and material hardship (Gershoff et al., 2007), maternal stress (Evans, Boxhill, & Pinkava, 2008), quality of the home environment (Krishnakumar & Black, 2002), and the physical home environment or cognitively stimulating materials and activities (Gershoff et al., 2007; Guo & Harris, 2000; Yeung et al., 2002).

Maternal psychosocial characteristics can affect the interaction between mother and child (Nair & Radhakrishnan, 2004; Petterson & Albers, 2001). Poor psychosocial functioning of the mother, possibly resulting from economic hardship, could lead to poor responsiveness (Evans et al., 2008) or poor provision of stimulation to the child (Baker-Henningham, Powell, Walker, & Grantham-McGregor, 2003), which in turn could negatively affect the child's cognitive outcome.

Social support refers to "any process through which social relationships might promote health and well-being" (Cohen, Gottlieb, & Underwood, 2000, p. 4) and is considered very important in the Indian context, where sharing of worries, living space, and food are common and valued practices (Pandey, 2006). Increased support can reduce parental stress, thereby positively affecting nurturing and parenting (McLoyd, 1990; Pascoe & Earp, 1984). Mother's social network has been linked to her responsiveness (Evans et al., 2008), stimulation in the home (Adamakos et al., 1986; Baker-Henningham et al., 2003), and structuring of the child's environment (Adamakos et al., 1986).

Economic hardship can lead to family conflict which in turn can add to disruptive parenting practices with negative consequences for children (Conger et al., 2002). Domestic violence is highly prevalent in India (International Institute for Population Sciences [IIPS] and ORC Macro, 2000). Studies from India also report a high prevalence of alcohol abuse in urban slums (D. Mohan, Chopra, Ray, & Sethi, 2000; D. Mohan, Desai, Chopra, & Sethi, 1992; I. Mohan, 1998), with particular impact on the family in terms of neglect and domestic violence (Gaunekar, Patel, & Rane, 2005). A relation between maternal alcohol use and the quality of the home environment was also found among African American children (Krishnakumar & Black, 2002).

Parenting can mediate between environmental factors and child outcome (Lugo-Gil & Tamis-LeMonda, 2008; McLoyd, 1998). Parents' use of cognitively stimulating materials and activities has been linked to the child's cognitive skills (Gershoff et al., 2007). Various versions of the Home Observation and Measurement of the Environment (HOME) Inventory (Caldwell & Bradley, 1984, 2003) correlated with children's academic performance in both Western and non-Western cultures (Bradley, Corwyn, & Whiteside-Mansell, 1996). Bradley and Corwyn (2005) indicated that three parenting aspects seem positively related to child development in every context, namely 1) warmth and responsiveness; 2) stimulation/teaching; 3) discipline/control. Parenting practices are influenced by socioeconomic background and culture; as a consequence, the specific parental behaviors that represent these three aspects can differ across and within contexts (Bradley & Corwyn, 2005; Harkness & Super, 2002).

Cross-cultural differences in warmth and responsiveness toward children can be found in behaviors such as praising, the public expression of affection, and spontaneous verbal interactions; for example, many Indian children are not allowed to speak without permission, resulting in few spontaneous verbal interactions (Bradley & Corwyn, 2005). Differences in stimulating and teaching involve both type and quantity. Many parents from Western, industrialized societies emphasize the importance of their children's cognitive development and the role of cognitively stimulating toys and materials (Bradley & Corwyn, 2005), whereas many Indian parents highly value their children's social skills, such as being helpful, respectful, and obedient (Saraswathi & Dutta, 1988; Sinha, 1988; Srivastava & Misra, 2001).

Low income families more often demonstrate an authoritarian parenting style and use physical disciplining to correct a child's behavior (Magnuson & Duncan, 2002). Studies from India report that parents are restrictive and hierarchical in their parenting and often resort to such physical disciplining (Hunter, Jain, Sadowski, & Sanhueza, 2000); however, the effects of this style on the (affective, cognitive, and behavioral) development of the child may not be negative, as is often assumed in Western countries (Chao & Tseng, 2002). A controversy exists over the effects of physical discipline on these child outcomes (Gershoff, 2002; Lansford et al., 2005). We do not refer here to harsh punishment and abuse, that will presumably have a universally negative relation with child outcome, but to mild disciplining. The inconsistency of findings on its effects could probably be resolved by taking into account the prevailing cultural norms about parental behavior. What could be seen as punishment and as having negative effects on child outcome by Western standards might be viewed as regular disciplining and as having a non-negative effect in India.

The present study

Figure 4.1 shows all the possible pathways from the described environmental factors to cognitive outcome. We expect the structure of the model to hold across cultures; yet, both the (non-) significance and strength of pathways may vary cross-culturally. By studying the applicability of our hypothesized model for our target sample of 6- to 10-year-old children of low SES in Bangalore (India), we can determine whether Western findings on environment-cognition relations generalize to this non-Western context, examine the differential roles of proximal and distal factors, and inform local interventions about critical environmental factors in an Indian context.

Method

Participants and study context

The sample consisted of 532 primary caregivers of at least one school-going child between 6 and 10 years old from Bangalore, South India (M = 8.70, SD = 1.17). Ninety-four percent of the interviewed caregivers were the mothers; the other six percent were grandmothers or other female relatives. Their mean age was 31.55 years (SD = 5.56, min. = 18, max. = 60).

Families had an average monthly income of 2700 Indian Rupees (56 USD). Many were illiterate (46% of the primary caregivers and 39% of the fathers) or had a maximum of five years of education (66% and 57%, respectively). Occupational levels were low: 65% of primary caregivers and 40% of fathers were unskilled workers (e.g., housemaid, helper, or manual laborer); 49% of fathers were skilled workers (e.g., carpenter, tailor, or painter). The father was absent in 13% of the families. The physical home environment was characterized by high levels of crowding; houses mostly had one or two rooms, and the average number of people in a household was 5.85 (SD = 1.89, min. = 2, max. = 19): 2.72 adults (SD = 1.22) and 3.13 (SD = 1.19) children. The main source of lighting was electricity. Most people made use of a public water tap and had their own toilet or shared a flush toilet with a few households. Children had very few toys to play with and had very limited access to books.

Instruments

Most of the instruments that were used in the current study needed adaptation to obtain a valid measure of the target constructs. Details on this process are given when necessary. After each instrument description, we report the internal consistency reliability (Cronbach's alpha) and the percentage of explained variance for the first factor extracted with principal component analyses. Evidence was found for the unidimensionality of all instruments. Some instruments employed response scales that were not the same for all items; therefore, scores were standardized for each item to compute total scores. For each (scale of each) instrument, the average z score across its items was used in further analyses.

Socioeconomic Status and Asset Indicators Questionnaire. The family's SES was derived from three questions addressing the highest level of education in the household (six categories, ranging from *illiterate* to *graduate*), the highest level of current occupation in the household (six categories, ranging from *unemployed* to *semi-professional*), and income, respectively. Correlations between these variables ranged from .11 to .25 ($p < .05$). Alpha was .40. The first factor explained 46% of the variance.

We assessed the family's housing conditions and goods (i.e., asset indicators) with one item on crowding (i.e., the number of rooms per person living in the house, reverse keyed) and five items from the Standard of Living Index (International Institute for Population Sciences (IIPS) and ORC Macro, 2000), namely the type of house (pucca, semi-pucca, or kutcha; pucca houses are completely made out of brick and concrete; semi-pucca have fewer of these materials; kutcha houses are made of mud and have thatched roofs), size of the house (kitchen in a separate room or not), type of fuel used for cooking, type of toilet facility, and the availability of household goods (such as electric fan and fridge). Correlations between these variables ranged from .15 to .41 ($p < .01$). Alpha was .39. The first factor explained 46% of the variance.

General Health Questionnaire. The General Health Questionnaire (Goldberg, 1978) is a measure of current mental health that has been widely applied in cross-cultural studies (Goldberg et al., 1997). The GHQ-12 is a shortened version of the General Health Questionnaire and its application as a screening tool in research settings is well documented. The GHQ-12 is commonly used in India to detect general psychological distress (see, for example, Pothen, Kuruvilla, Philip, Joseph, & Jacob, 2003); an existing Kannada translation of the instrument (Sriram, Chandrashekar, Isaac, & Shanmugham, 1989) was used in the present study as a measure of psychological functioning of the primary caregiver. An example of an item is "Have you recently lost much sleep over worry?". There were four answer options (0 = *not at all*; 1 = *no more than usual*; 2 = *rather more than usual*; 3 = *much more than usual*). Scores were recoded so that higher scores indicated a higher level of general mental health. Alpha was .82. The first factor explained 35% of the variance.

Social Support Questionnaire. The items of our measure of social support were based on and adapted from existing questionnaires and example items given in articles and book chapters (see, for example, Wills & Shinar, 2000). Social support takes different forms, namely emotional, instrumental, informational, and appraisal (Tardy, 1985). Our final version resulted from an intensive procedure of translating, piloting, and adapting (see Procedure), and consisted of 12 items: 6 reflected informational or instrumental support (e.g., "Do you feel there are enough people in your environment that would lend or give you something you need, like food, clothing, or money?"), and 6 reflected emotional support or appraisal (e.g., "Do you feel there are enough people that can comfort you when you feel unhappy about your daily life?"), with answer options ranging from *definitely not enough* (1) to *definitely enough* (4). Alpha was .88. The first factor explained 43% of the variance.

Family Conflict Questionnaire. Three items were selected from the Kannada version of the Family Environment Scale (Moos, Insel, & Humphrey, 1974) to assess household risk factors. The items were: "Are there a lot of fights (i.e., arguments) in your family?", "Do family members sometimes get so angry they throw things?", and "Do family members sometimes hit each other?" Each question could be answered by *yes* (= 1) or *no* (= 0). Given the high prevalence of alcohol abuse and related domestic violence, we added an item on alcohol (ab)use with two dependent parts: "Does anybody in the family drink alcohol?", (if *yes*:) "Does this cause any disturbance at home (e.g., shouting, violence)?". These two parts were combined into one single item with three answer options (i.e., *no alcohol use*; *alcohol use but this does not cause disturbance*; *alcohol use which causes disturbance*). Alpha was .79. The first factor explained 62% of the variance.

Middle Childhood HOME Inventory. The Home Observation for Measurement of the Environment (HOME) Inventory (Caldwell & Bradley, 1984, 2003) was developed to measure the quality and quantity of stimulation and support available to a child in the home environment. We used the Middle

Childhood HOME (59 items), designed for children between 6 and 10 years old. Because our study sample was of low SES and also differed substantially in culture from the American norm sample, the instrument needed extensive adaptation. Some items were not appropriate given the low SES context, such as “Family member has taken child to a scientific, historical, or art museum within the past year”. Other items showed no variance in our relatively homogeneous sample, such as “The interior of the home is not dark or perceptually monotonous”. These items were removed from the list. Some items were adapted so as to introduce variability and avoid floor effects; for example, we changed the item “Child has free access to at least ten appropriate books” to “Child has free access to children’s books”. Also, some items were added that were deemed appropriate in our Indian sample and could be related to developmental outcomes, such as items related to disciplining the child. The original version contains one item on physical punishment. Because disciplining the child is very common in our sample and can take different forms, we elaborated on this topic by using the question “How do you discipline your child?”. Parents had to indicate whether or not and if so, how many times in the past two weeks, they applied 1) verbal disciplining (shouting, scolding); 2) verbal disciplining (discussions); 3) physical disciplining. Finally, the items in which the interaction between caregiver and child is directly observed were not used. We collected our data during the day, and in most of the cases the children were at school, making a direct observation of the child-caregiver interaction impossible. Our adapted instrument consisted of 24 items.

In line with Krishnakumar and Black’s (2002) suggestion, we aimed to distinguish between clusters of parenting behaviors that might have differential effects on cognitive performance. However, since the adaptation only contained 24 items, the eight subscales of the original HOME were not properly covered. The dimensions of warmth, discipline, and stimulation that Bradley and Corwyn (2005) described could not be fully covered because the category of warmth was not well-presented due to the lack of items that required direct caregiver-child interaction. A Disciplining subscale of four items was made; an example of an item is "How many times in the last two weeks have you verbally disciplined your child by shouting or scolding?". For the remaining 20 HOME items that reflect stimulation, we applied a distinction that is in line with our conceptual model of distal and proximal variables (see Figure 4.1); the items were split into a proximal subscale of Direct Involvement subscale and a distal subscale of Indirect Involvement.

The Direct Involvement subscale contained 10 proximal items reflecting direct involvement with the child, by providing structure (e.g., "Does your family have a regular and predictable program every day for the child?") and by providing direct stimulation (e.g., "Do you encourage your child to read on his/her own?"). The Indirect Involvement subscale consisted of 10 (more distal) items reflecting more indirect involvement with the child, by providing an enriching climate. There were items related to family companionship and items related to activities with the children (e.g., "How often has the family member taken the child to some type of live performance or a celebration within the past year?"), and items related to indirect cognitive stimulation (e.g., "On how many days a week does the child have contact with friends outside school hours?", and "On how many days a week does one or do both parents watch/hear/read the news on television/radio/newspaper?").

Most items could be answered by *yes* (= 1) or *no* (= 0). If necessary, items were recoded so that a higher score reflected more stimulation. For some items the answer was a frequency (e.g., "How many times a month does the family visit relatives or friends or do relatives or friends visit the family?"). A cut-off point was determined based on the variance in the data; all frequency items were dichotomized, distinguishing a frequency of zero (recoded into *no* = 0) from higher frequencies (recoded into *yes* = 1). Alpha values were .54 for the Disciplining scale, .48 for the Direct Involvement scale, and .48 for the Indirect Involvement scale. The first factor explained 48%, 19%, and 22% of the variance, respectively.

Kaufman Assessment Battery for Children, Second Edition (KABC-II). The KABC-II is an individually administered measure of cognitive ability that can be used for children from 3 to 18 years of age (Kaufman & Kaufman, 2004). Eight KABC-II subtests were selected and three other subtests were added. The adaptations and validation procedure of the adapted subtests are described elsewhere (Chapter 2; Chapter 3). A general cognitive score was calculated by averaging the age-standardized sum scores for each of the four scales underlying the cognitive battery (i.e., fluid reasoning, short-term memory, long-term storage and retrieval ability, and cognitive speediness). This score was used in the present study as an index of the child's cognitive performance. Alpha was .86. The first factor explained 61% of the variance.

Procedure

Adaptation and pilot phase. Kannada translations of the GHQ-12 and of the items composing our Family Conflict Questionnaire were already available. A pilot study confirmed their ecological appropriateness. To ensure the appropriateness of the constructed Social Support Questionnaire and of the adapted HOME Inventory, we applied an iterative procedure of translating, piloting, and adapting the test items. All items were first translated by a team of psychologists and independently back translated by a psychologist not involved in the project. Linguistic and cultural appropriateness were checked by performing a pilot study, in which the interviewed caregivers provided feedback on the clarity and relevance of the items, and could give suggestions for inclusion of additional items. Changes were implemented if necessary, the translation was adjusted accordingly, and the new version was piloted again until an appropriate instrument was obtained. Eighteen caregivers were interviewed to obtain the final version.

Data collection phase. Two research assistants interviewed the primary caregivers at their children's school to record demographic and socioeconomic information. Two social workers who were fluent in both English and Kannada were intensively trained by a local psychologist (the fifth author) to administer the other questionnaires to the primary caregivers. The trainer joined each social worker for 10 interviews at the start of data collection to determine the inter-rater reliability. Average intraclass correlations (absolute agreement) between the trainer's and the social workers' scoring were very high: .887 for the GHQ-12, .957 for the Social Support Questionnaire, 1.000 for the Family Conflict Questionnaire, and .996 for the HOME Inventory. On average, each social worker then administered the questionnaires to three caregivers a day at their homes and each interview session took about 45 minutes. More than 95% of the interviews were administered in Kannada, the remainder in Tamil.

Results

First, some findings on the more proximal variables in the home environment are described. Second, the link between the home environment and children's cognitive performance is analyzed by structural equation modeling (Arbuckle, 2005).

Descriptives for proximal variables

We take a closer look at the more proximal variables reflecting the family climate (i.e., general mental health, social support, and family conflict) and parental behaviors. To give an indication of the general mental health of the primary caregivers, their average item score on the GHQ-12 was examined. This score was 1.93 on a scale of 0 to 3, indicating an on average good state of self-reported mental health. Unlike what we would expect in a low SES urban sample, the score distribution of GHQ-12 did not show any overrepresentation of caregivers with major psychological distress; on the contrary, the distribution was negatively skewed (z = -9.41, $p < .01$). The average item score on the Social Support Questionnaire was 2.73 on a scale of 1 to 4, indicating sufficient perceived social support. Family conflicts were common; about half of the primary caregivers reported many verbal arguments in their families, 40% reported incidents of physical violence. Half of the primary caregivers reported alcohol use of a family member, and in 70% of those cases this caused disturbance at home, such as shouting or violence.

An analysis of the HOME Inventory items showed that about 75% of primary caregivers reported praising their child and 80% encouraged their child to read; yet, about 90% of the children did not have access to books besides their school books, and many parents could not read with them because they were illiterate. Almost half of the children did not have contact with friends after school hours. More than half of the parents took their child to a live performance or celebration at least once a year (on most occasions the event was a wedding). Around 90% of primary caregivers reported regularly losing temper with their children, 74% reported shouting at them, 66% used verbal discussions and around 80% used physical disciplining. The HOME Inventory showed that caregivers were highly involved, paying both positive and negative attention to their children.

Home environment and cognitive performance

Table 4.1 presents the bivariate correlations between all questionnaires and cognitive performance. SES and asset indicators significantly related to all variables except the disciplining items from the HOME, emphasizing the broad network of associations of SES. Family climate variables (i.e., general mental health, social support, and family conflict) related to all parental behaviors (direct involvement, indirect involvement, and disciplining), with the exception of general mental health, which was not related to the direct involvement with the child. Disciplining the child was related to the family climate variables but not to the other parental behaviors of direct and indirect involvement and not to cognitive performance. Direct involvement with the child was the only proximal variable that correlated significantly with cognitive outcome.

Path analysis was used to test the fit of the model displayed in Figure 4.1 to the data. All possible relations between all variables were included, except for one so as to have one degree of freedom to estimate the model fit (the direct relation

Table 4.1

Bivariate correlations among all questionnaires and cognitive performance

	SES	AI	GHQ	SS	FC	H-D	H-I	H-Dis	CP
SES	.67								
Asset Indicators	.36**	.62							
GHQ-12	.09*	.15**	.58						
Social Support	.15**	.18**	.37**	.65					
Family Conflict	-.09*	-.21**	-.27**	-.22**	.78				
HOME-Direct	.11**	.15**	.06	.13**	-.14**	.42			
HOME-Indirect	.23**	.28**	.19**	.23**	-.19**	.21**	.42		
HOME-Disciplining	.02	-.03	-.11**	-.11*	.18**	-.01	-.00	.69	
Cognitive Performance	.11*	.09*	.07	.01	-.02	.21**	.07	.01	.65

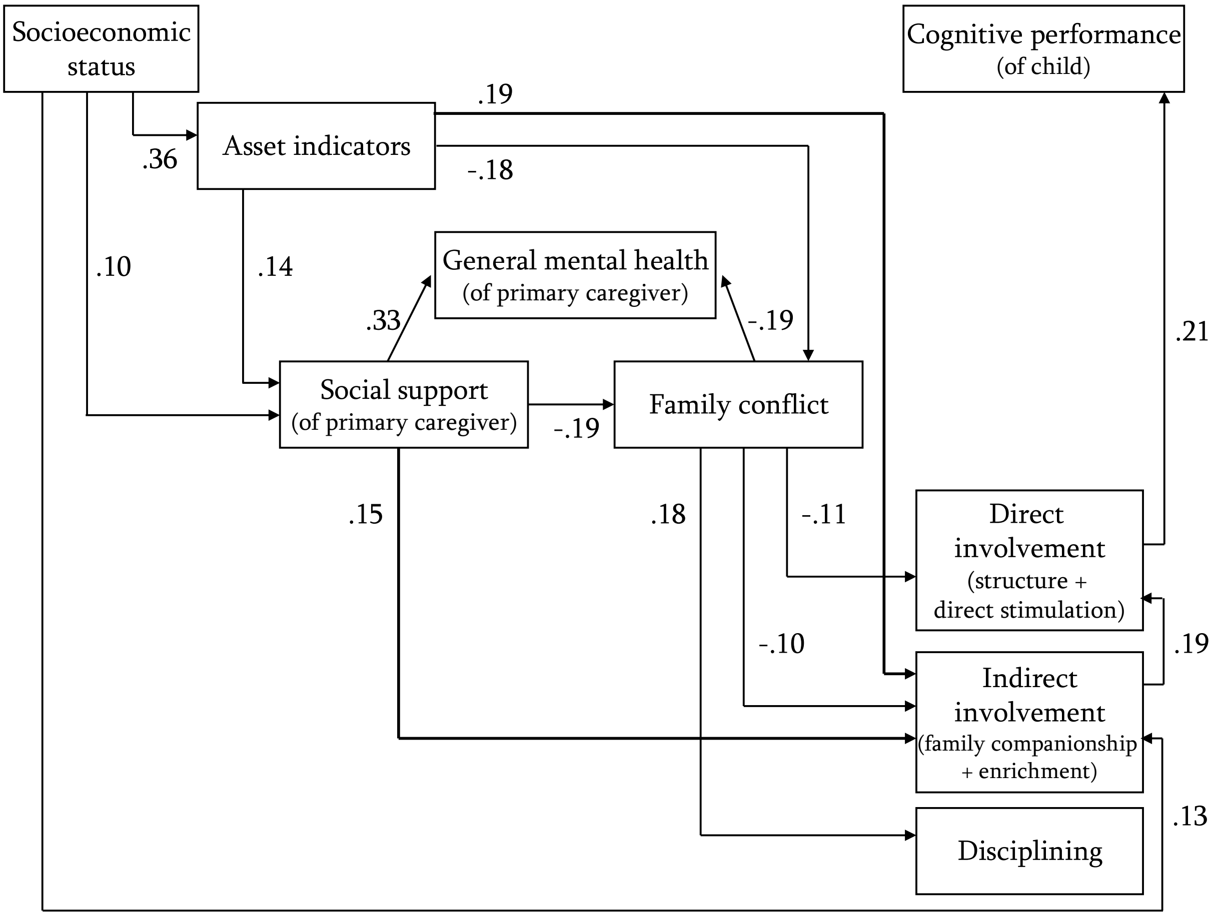
Note. SES = Socioeconomic Status, AI = Asset Indicators, GHQ = General Health Questionnaire, SS = Social Support, FC = Family Conflict, H-D = HOME-Direct, H-I = HOME-Indirect, H-Dis = HOME-Disciplining, HOME = Home Observation for Measurement of the Environment, CP = Cognitive Performance.

Standard deviations are given on the diagonal.

* $p < .05$. ** $p < .01$.

between SES and the Disciplining scale was not included; we did not expect income and educational or occupational level to be directly related to disciplining the child). The model had an excellent fit (χ^2[1, N = 532] = 1.23, p = .27, χ^2/df = 1.23, GFI = 1.00, AGFI = .98, TLI = .98, CFI = 1.00, RMSEA = .02, AIC = 89.23); however, many pathways were not significant. Figure 4.2 displays the model in which only the significant pathways were retained. The fit (χ^2[21, N = 532] = 26.95, p = .17, χ^2/df = 1.28, GFI = .99, AGFI = .98, TLI = .97, CFI = .98, RMSEA = .02, AIC = 74.95) was not significantly different from the fit of the first model ($\Delta\chi^2$[20, N = 532] = 25.72, p = .18). The model in Figure 4.2 matched the hypothesized model in that the relations between the distal environmental variables (SES and asset indicators) and cognitive outcome were mediated by the proximal environmental factors (family climate and parental behaviors). SES was only directly related to asset indicators, social support and the indirect involvement with the child. SES was only related to the child's cognitive outcome through the family's asset indicators, the family climate and the involvement with the child. The general mental health of the primary caregiver did not show any direct association with parental behaviors; family climate was related to the parents' indirect involvement with the child through perceived social support and family conflict and to the parents' direct involvement through family conflict. Only the most proximal scale, the direct involvement of parents with their children through structure and stimulation, was directly and significantly related to the cognitive score. All indirect effects were significant ($p < .05$).

The percentages of explained variance for all variables in this model show that the effects were small to medium despite being significant: small for social support (4%), family conflict (8%), direct involvement (5%), and disciplining (3%), and medium for asset indicators (13%), general mental health (17%), and indirect involvement (13%). The entire model explained a modest 4% of the variance in cognitive performance.



Note. Standardized coefficients. Cognitive performance is standardized for age.
All depicted parameter estimates are significant, $p < .05$.

Figure 4.2
The model relating home environment to cognitive performance in Bangalore, India

Discussion

Most studies describing the relations between children's home environment and cognitive outcomes have been done in the United States. We were interested in studying the generalizability of findings on these relations to a non-Western setting and examined which environmental variables contributed to the cognitive performance of 6- to 10-year-old school-going children of low SES in Bangalore, South India. Our path model confirmed the environment-cognition relations as suggested by the investment and family stress model. The hypothesized distinction between distal and proximal variables was confirmed in that more distal variables were at least partly linked to cognitive outcomes through more proximal variables. In line with the investment model, SES (an indicator of potential investment in the child) and asset indicators (i.e., housing conditions and goods, which are concrete indicators of investment) were associated with cognitive outcome, but merely through parental behaviors, notably the parents' direct involvement with the child (i.e., the most proximal parental behaviors, such as applying rules, encouraging the child to read, and encouraging self-care routines). This means that having the financial resources to invest in living conditions and household goods and to invest in cognitively stimulating materials and activities is related to child outcome. The family stress model was supported by showing that the family climate (i.e., general mental health and perceived social support of the primary caregiver, and family conflict) was related to the asset indicators; the stress accompanying poverty and the availability of few goods could negatively affect the family climate. The associations between family climate and a child's cognitive performance were fully mediated by the parents' direct involvement with the child. Experimental or longitudinal studies would be needed to confirm the causal nature of the relations in Figure 4.2.

What can be concluded about the adequacy of the proximal-distal distinction underlying both the investment and the family stress model as presented in Figure 4.1? First, the findings confirmed the hypothesized structure of the relations between the variables. Second, the proximal-distal distinction proved useful both to describe how SES is related to cognitive outcome via mediating variables and to split up the items of the HOME Inventory. Differential relations with cognitive performance were found for the more distal and more proximal aspects of parental behaviors. The correlation between the proximal Direct Involvement subscale and cognitive performance (.21) was significant, whereas the correlation of the more distal Indirect Involvement subscale (.07) was nonsignificant. The correlation between the (proximal) Disciplining subscale (.01) and outcome was, however, also nonsignificant. Disciplining usually has stronger relations with a child's social-emotional than cognitive functioning (Paolucci & Violato, 2004). The correlation between the HOME Inventory as a whole (without distinguishing subscales) and cognitive performance was .13. Splitting the scale into distal and proximal subscales provided us with more specific information as to which parental behaviors contributed more to cognitive outcome. This differential pattern of correlations provides additional support for the relevance of the distinction between proximal and distal aspects in environment-cognition relations.

The last conclusion involves the salience of the statistical associations observed, by examining 1) the significance of pathways, 2) the strength of significant pathways, and 3) the percentage(s) of explained variance. First, many pathways of the conceptual framework as depicted in Figure 4.1 were not significant in our sample. As explained in the introduction, Figure 4.1 shows all possible pathways between the environmental variables and cognitive outcome; different pathways might be significant for different samples. With our sample we found proximal variables fully mediating the association between distal variables and outcome; hardly any factors in the home environment of the child were directly associated with cognitive performance. This finding implicates that the association between SES and the child's cognitive outcome, which was significant, could be successfully "unpackaged" by the mediating variables; SES is not the critical variable in this sample in determining cognitive outcome, but SES can be psychologically interpreted as a proxy for psychological processes, such as direct parental involvement, that are related to cognitive outcome. Second, the significant correlations and standardized regression coefficients in our model were generally equal to the values reported in literature on Western sample, however, some were lower. For example, beta values for the relation between SES variables and cognitively stimulating materials or activities range from .13 to .51 in the literature (Gershoff et al., 2007; Guo & Harris, 2000; Yeung et al., 2002). Our values were at the lower end of this range. The same holds for the relation between cognitive stimulating materials or activities and cognitive skills, with correlations in the literature ranging from .20 to .60 (see Bradley et al., 1996). Also, standardized regression coefficients of indirect associations between SES and cognitive outcome ranged from .02 to .30 in previous research (Gershoff et al., 2007; Guo & Harris, 2000; Krishnakumar & Black, 2002; Yeung et al., 2002). Third, with just one (significant) direct and many indirect pathways between home environment and cognitive performance, only 4% of the variance in cognitive performance could be explained. Divergent findings have been reported in the literature, with values roughly ranging from 4 to 40% (Bradley et al., 1996; Gershoff et al., 2007; Guo & Harris, 2000; Yeung et al., 2002).

The low values of some of the regression coefficients and the modest amount of variance explained in cognitive outcomes deserve closer scrutiny. More specifically, the question should be addressed whether these values are a valid reflection of the associations in this cultural context or whether they are due to limitations of our study. The first reason for some weak associations is the relative homogeneity of our sample. Involving families with a higher SES would presumably increase the amount of variance explained. Second, the explained variance might have also increased by including other child variables such as nutritional status and health (Walker et al., 2007). Also, of the three microsystems affecting a school-age child (i.e., home, school, and peers; Bronfenbrenner, 1979), we have now only included the home environment.

A third factor contributing to the low values could be the low internal consistency reliabilities of some instruments. Questionnaires to assess SES and asset indicators were short, possibly explaining their low Cronbach's alpha values. The low internal consistencies of the adapted HOME Inventory subscales might be due to a combination of the homogeneity of our sample and the relatively small numbers of items in each scale. Also, parents obtained different patterns of scores on the Direct Involvement subscale; some parents scored high on certain items while other parents scored high on others. This implies that it does not matter what kind of stimulation is given to obtain a positive relation with cognitive outcome, as long as there is sufficient stimulation. The dissimilar patterns of scores across caregivers could have contributed to the low alpha value.

The possible low content validity of the adapted HOME Inventory is a fourth factor; did the HOME adaptation actually cover the parental behaviors that are important for children's cognitive outcome in an Indian context? Our adaptation was carefully made, together with local informants, and we argue that it is doubtful whether a further adaptation would increase the reliability and content validity, and thereby the relation with cognitive performance. The internal consistency and ecological validity of our measures might be boosted by interviewing multiple caregivers instead of merely the primary caregiver, given that many Indian children live in extended families where responsibilities for child care are shared.

Fifth, the relation between the home environment and developmental outcomes is usually stronger for younger children than for children in middle childhood, because the latter are exposed to a greater degree of external influences (e.g., school, peers) than merely the home environment (Bradley et al., 1996). Many of the studies that report correlations between the HOME and cognitive outcome as high as .60 deal with infants or children in early childhood. Sixth, the primary caregivers in the present study might view cognitive stimulation as part of the (cognitive) education that takes place at school; perhaps they do not see themselves as playing a role in the cognitive development of their child. Parents of low SES generally believe to have less control over their children's development than parents of higher SES (Hoff, Laursen, & Tardif, 2002). Last, parents' expectations might play a role. The lower the SES, the higher Indian parents tend to value children's social skills as compared to cognitive skills (Srivastava & Misra, 2001).

It can be concluded that by increasing the internal consistency of some measures, the amount of variance in cognitive outcome that is explained by our predictors might have increased somewhat; yet, there are also strong reasons to believe that the link between SES, parenting, and cognitive outcome is not very strong in this Indian context and adding new samples or measures would not have altered the modest role of the variables in the present study.

The primary caregivers in the present study generally reported an unexpectedly good state of mental health, as manifested in the relatively high mean score level on the GHQ-12. The question formulation of the GHQ might have played a part in this finding. The instrument asks about certain feelings or experiences relative to "usual" feelings or experiences. As a result, the GHQ is more sensitive to detect short-term psychological distress than to detect chronic distress (Goodchild & Duncan-Jones, 1985). For example, when people are already under constant strain (as could be the case in our sample of low SES), they could answer the question "Have you recently felt under strain?" with *no more than usual*. This would be an indication of good mental health according to the GHQ scoring, but on an absolute level, these people could be under significant strain. The GHQ might not provide adequate information on the interviewee's baseline level of mental health. An inaccurate representation of the general mental health of the primary caregiver could have caused the absence of a direct relation with parental behaviors in the path model of the present study.

Even with a homogeneous sample we have found support for the relative importance of proximal and distal variables. How can the present study inform interventions aimed at fostering child development? The relations between SES and outcomes are mostly indirect and it will probably take a long time for positive changes in SES (by educating caregivers or by an increase in family income) to bring about positive changes in outcome. Our study suggests that influencing parental behavior towards greater involvement with the child is critical for cognitive performance. It is much more (cost) effective and feasible on a short-term to increase caregivers' awareness of how they themselves can directly affect developmental outcomes by being involved with their child through providing structure and stimulation.

Chapter 5

Rugby Versus Soccer in South Africa: Content Familiarity Explains Most Cross-Cultural Differences in Cognitive Test Scores[4]

Cross-cultural differences in cognitive test scores are not well understood. Where do they come from and why are they larger for some tests than for others? Spearman's Hypothesis (SH) relates these cross-cultural differences to the cognitive complexity of tests; differences are larger for tests with a higher cognitive complexity (Jensen, 1985, 1998). SH attributes this pattern to cross-cultural differences in the underlying general cognitive ability on which tests with a higher cognitive complexity more strongly rely. However, a test's cognitive complexity is often confounded with its cultural complexity, and the latter may largely explain cross-cultural score differences (Helms-Lorenz, Van de Vijver, & Poortinga, 2003). A test's cultural complexity (referred to as "cultural loading" in Chapter 2) refers to the specific cultural knowledge that is required to perform well on this test, such as declarative and procedural knowledge that is shared in a particular culture. The cultural complexity of a test is reflected in a group's familiarity with the type of test and with the content of the test. In order to disentangle the influence of cultural and cognitive complexity on test performance, the present study examines the effect of content familiarity on the performance on tests of different cognitive complexity. More specifically, we address the role of content familiarity in tests measuring (the cognitively complex ability of) fluid reasoning and tests measuring (the less complex, though related abilities of) short-term memory, attention, and working memory (Carroll, 1993; McGrew, 2005). For each of these tests, two test versions were developed in which the content familiarity of the items was maximized for either Afrikaans-speaking or Setswana-speaking school-age children in South Africa. Both versions were administered to children of both groups.

---

[4] This chapter is based on Malda, Van de Vijver, & Temane (2009)

Cultural complexity

Successful performance on a test with a high cultural complexity requires specific cultural knowledge. This knowledge is stored in a semantic network in memory, which can be viewed as a system of nodes (cultural elements) with links (associations) between them. This network is comparable to the system of cognitive elements required for cognitive skill acquisition (see Anderson, 1982). For people who are familiar with the culture in which a test is developed, this semantic network has a well defined structure of strong and weak links, which means that relevant associations are readily made between the content of the test and their knowledge. This facilitates successful completion of the test. People from a different culture do not have this well developed semantic network associated with the content of this particular test, because they may not know the cultural elements or their associations; as a consequence, they have difficulty to perform well. The level of cultural complexity of a test then refers to the extent to which an elaborate and automated semantic network of cultural information is required to perform well. Cultural complexity is conceptualized in the present study as the extent to which test content (i.e., words, drawings) is more familiar to one of the compared groups.

Cognitive abilities in the present study

Short-term memory is described as “the ability to apprehend and maintain awareness of elements of information in the immediate situation” (McGrew, 2005, p. 153). Controlled attention is defined as “the capacity to maintain and hold relevant information in the face of interference or distraction” (Swanson, 2008, p. 582). Working memory is “a system for the simultaneous processing and storage of information” (Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000, p. 1018). The assignment of attention to the contents of short-term memory creates working memory (Schweizer & Moosbrugger, 2004; Swanson, 2008). Fluid reasoning is defined as “the use of deliberate and controlled mental operations to solve novel, ‘on-the-spot’ problems (i.e., tasks that cannot be performed automatically)” (McGrew, 2005, p. 151).

Some researchers state that short-term memory and working memory cannot be differentiated in children (Hutton & Towse, 2001); however, others have shown that they are already distinguishable from six years of age (Gathercole, Pickering, Ambridge, & Wearing, 2004; Swanson, 2008). Working memory capacity and fluid reasoning are strongly related (Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002), though distinguishable (Ackerman, Beier, & Boyle, 2005). Conway, Cowan, Bunting, Therriault, and Minkoff (2002) indicated that rather complex tasks such as working memory tasks do not rely on automated routines, similar to fluid reasoning tasks. Working memory and reasoning tasks share a demand for controlled attention. Both working memory and attention play a role in fluid reasoning (Unsworth & Engle, 2005); they have a direct relationship with fluid reasoning and the relation between attention and fluid reasoning is mediated by working memory as well (Schweizer & Moosbrugger, 2004). The cognitive structure underlying fluid reasoning abilities that is compatible with these findings is shown in Figure 5.1. Short-term memory and attention have both direct and indirect relations with fluid reasoning. Working memory plays a mediating role. Going from left to right in Figure 5.1, the cognitive abilities become cognitively more complex.

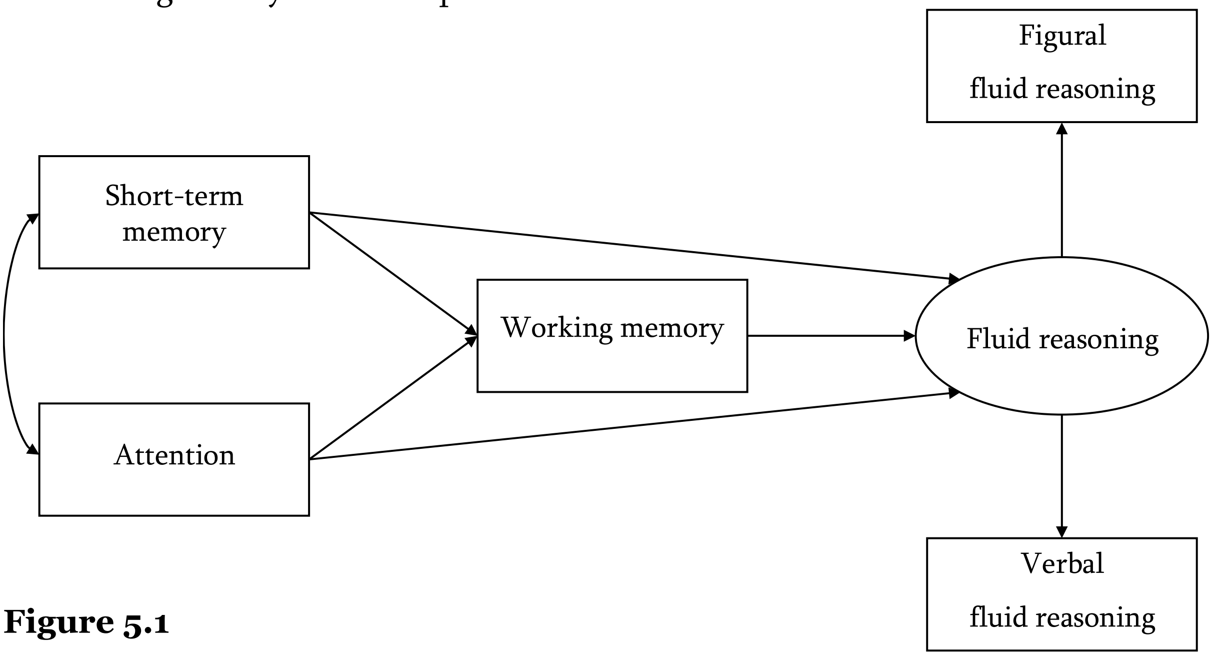


Figure 5.1
Hypothesized cognitive structure

Cognitive versus cultural complexity

Only small cross-cultural differences have been found in attention and short-term memory, larger differences in working memory, and the largest differences have been reported in fluid reasoning. How can this patterning be explained? One explanation, known as Spearman's Hypothesis (SH), holds that tasks with a higher cognitive complexity show larger cross-cultural score differences, mainly because of assumed cross-cultural differences in the underlying general cognitive ability on which high cognitively complex tests strongly rely (Jensen, 1985, 1998). Fluid reasoning tasks produce the largest cross-cultural differences because of their large cognitive complexity when compared to attention, short-term memory, and working memory tasks (Carroll, 1993). Jensen has conducted many studies that supported SH (reviews can be found in Jensen, 1985, 1998). A number of other studies also found support for SH (e.g., Hartmann, Kruuse, & Nyborg, 2007; Lynn & Owen, 1994; Te Nijenhuis & Van der Flier, 1997).

SH has met with both statistical and conceptual criticism. Statistical analyses employed to test the hypothesis have been questioned. Multigroup confirmatory factor analysis has been proposed as a statistically more rigorous procedure for testing SH than Jensen's method of correlated vectors; a re-analysis of two data sets that were supportive of SH when analyzed by Jensen's method failed to meet basic requirements of cross-cultural comparability in a multigroup confirmatory factor analysis (see Dolan, Roorda, & Wicherts, 2004). There are substantive reasons to doubt the importance of cognitive complexity in explaining cross-cultural score differences. Affluence and socioeconomic status have been found to explain a large part of cross-cultural score differences through educational differences (Van de Vijver, 1997). The content of a test or the medium in which a test is administered has also been shown to determine how well a child performs. Serpell (1979) tested the perceptual skills of British and Zambian children by administering similar tasks in different media. British children performed better on paper and pencil tasks whereas the Zambian children performed better on wire-modelling tasks. Carraher, Carraher, and Schliemann (1985) found that

Brazilian school-going children performed better on arithmetic tasks when they were presented in the form of a problem (as in an everyday market situation) than when they were presented as numerical calculations (as in school). The explanation of cross-cultural score differences that we test in the current study holds that cognitive complexity is usually confounded with cultural complexity, and that the latter is the actual factor explaining most of the cross-cultural score differences (Helms-Lorenz et al., 2003). We label this explanation the Cultural Complexity Hypothesis (CCH). Tests that address simple information processing usually show less cultural bias than tests addressing complex information processing (Vock & Holling, 2008). The former tests, measuring abilities such as attention and short-term memory, do not employ complex cultural information and hence, they are not very sensitive to group (and individual) differences in access to cultural information. Cross-cultural differences on these tests are expected to be small. The differences are larger on cognitively more complex processes, such as working memory, and will be largest on the most complex tasks, such as fluid reasoning tasks, which often require extensive cultural information to solve them.

Present study and hypotheses

A test was constructed for each of the abilities of Figure 5.1. There were two versions of each test. One version contained items with a relatively higher content familiarity for the Afrikaans-speaking ("Afrikaans") than for the Setswana-speaking ("Tswana") children, labeled the Afrikaans test version. The other contained items with a relatively higher content familiarity for the Tswana than for the Afrikaans children, labeled the Tswana version. The study involved both urban and rural Tswana children, who differed in their exposure to Afrikaans culture (with urban children being more exposed than rural children) and to more traditional Tswana culture (with rural children being more exposed than urban children). Differences in test scores between the Afrikaans and Tswana children are expected to be due to the content familiarity of the tests. Four groups of hypotheses are tested. The first group involves the overall patterning of the results, as can be derived from both CCH and SH:

1. The underlying cognitive structure as displayed in Figure 5.1 is a) valid for the overall sample; b) invariant across test versions (i.e., Afrikaans and Tswana versions); c) invariant across groups (i.e., Afrikaans, urban Tswana, and rural Tswana children).
2. Going from left to right in Figure 5.1, score differences (between groups for each test version and between test versions for each group) are expected to increase. Small or no group and version differences are expected for attention and short-term memory, larger differences are expected for working memory, and the largest for fluid reasoning.

The second group of hypotheses predicts that children's performance on the test version designed for their own culture is at least as high as on the test version designed for the other culture. These hypotheses are necessary for CCH to be correct and not necessary for (but in accordance with) SH:

3. Afrikaans children score at least as high on the Afrikaans test version as on the Tswana version (more specifically, differences between test versions are absent or small for short-term memory and attention, larger for working memory, and largest for fluid reasoning).
4. Rural Tswana children score at least as high on the Tswana version as on the Afrikaans version (differences between test versions are absent or small for short-term memory and attention, larger for working memory, and largest for fluid reasoning).
5. Urban Tswana children score similar on the Afrikaans and Tswana test version.

The next group of hypotheses is critical for testing whether CCH or SH is supported. CCH predicts that it depends on the test version which group obtains the highest score and that one group scores at least as high as the other group(s) on the test version developed for its own culture. SH predicts that one group scores consistently higher than the other(s) on both test versions (apart from random fluctuations). The following hypotheses are tested:

6. Afrikaans children score at least as high as urban and rural Tswana children on the Afrikaans test version (differences between groups are absent or small for short-term memory and attention, larger for working memory, and largest for fluid reasoning).
7. Rural Tswana children score at least as high as Afrikaans children on the Tswana test version (differences between groups are absent or small for short-term memory and attention, larger for working memory, and largest for fluid reasoning).

Both hypotheses need to be confirmed to support CCH. If hypothesis 6 (or 7) is confirmed for the working memory and fluid reasoning tests, SH would predict that 7 (or 6) is automatically disconfirmed.

The final hypotheses involve the relative positions of the scores of the urban Tswana children in between the scores of the two other groups. Confirmation of these hypotheses would be in line with both CCH and SH:

8. Urban Tswana children score in between Afrikaans and rural Tswana children on the Afrikaans test version.
9. Urban Tswana children score in between Afrikaans and rural Tswana children on the Tswana test version.

Method

Participants

The sample consisted of 501 South African primary school children (245 girls, 256 boys) from grades 3 and 4, with an average age of 9.37 years (*SD* = 1.05). One hundred sixty-one were white urban Afrikaans-speaking children from two primary schools in the town of Potchefstroom, North West Province; 181 were black urban Setswana-speaking children from two primary schools in Ikageng, a township near Potchefstroom; 159 were black rural Setswana-speaking children from three primary schools in Ramatlabama, a rural setting 15 kilometres outside of the city of Mafikeng, North-West Province. Besides the linguistic distinction between the subsamples (Afrikaans versus Setswana), there was a cultural distinction (Afrikaans versus Tswana) and an urban-rural distinction. The latter distinction, however, was relative in that the children from Potchefstroom ("Afrikaans") and Ikageng ("urban Tswana") lived in an urban area as compared to the more rural area of Ramatlabama ("rural Tswana").

Most of the houses of Afrikaans children were made of bricks and had tiled roofs, while most houses of the Tswana children had walls of either corrugated iron or bricks and roofs of corrugated iron. Eighty-one percent of Afrikaans children had their own room as opposed to 31% of the urban Tswana and 40% of the rural Tswana children. Afrikaans children had 2.36 cars per family on average, whereas 35% of the urban Tswana families and 50% of the rural Tswana families had a car.

Instruments

Five cognitive tests were constructed: a short-term memory test, an attention test, a working memory test, a figural fluid reasoning test, and a verbal fluid reasoning test. There were two cultural versions of each test, based on the relative familiarity of item content: an Afrikaans and a Tswana version. The two test versions were developed in a three month pilot phase. We visited children's homes and schools, and spoke to parents, teachers, and specialists (e.g., a child psychologist, speech therapist) to obtain information regarding words, objects, customs and practices that were familiar to the Afrikaans and Tswana children, respectively. Pilot testing took place at three schools and involved 50 children. An iterative procedure was applied of translating the test instructions into the local languages, administering the instruments to a small number of children, and adapting the content and/or instructions if necessary, until the instruments were deemed appropriate. The instructions and items of both test versions were developed in English and then translated into Afrikaans and Setswana. Afrikaans children did (Afrikaans and Tswana) tests in Afrikaans; Tswana children did (Afrikaans and Tswana) tests in Setswana. The short-term memory test, attention test, and working memory test had a discontinuation rule: these tests were stopped after the child failed three consecutive items. The figural and verbal fluid reasoning tests did not have a discontinuation rule.

Short-term memory test. This individually administered test consisted of 24 items and required the child to repeat word sequences, varying from two to nine words, read out loud by the test examiner. Both the Afrikaans and Tswana test version used meaningful words with a higher familiarity for the Afrikaans and Tswana children, respectively. Examples of words used in the Afrikaans test version are “computer”, “camera” and “shower”, and examples of the Tswana test version are “tuckshop” (a small food shop that is common in the Tswana community), “soccer”, and “braids”.

Attention test. The child’s task in the individually administered attention test was to count the number of times he/she heard a pre-specified group of words in the sequence of words read out loud by the test examiner. The test consisted of 24 items. The Afrikaans version targeted groups of two electrical appliances followed by one piece of clothing (e.g., Heater – Iron – Trousers); the Tswana version aimed at groups of two family members followed by one animal (e.g., Aunt – Son – Dog).

Working memory test. The first 3 items of this individually administered test required the child to judge whether a statement was true or false. The following 18 items required the child to judge whether a statement was true or false and remember this while listening to statements that followed. After the test examiner finished reading all statements in a single item, the child was asked to say for each of them whether it was true or false, in the same order as the examiner read them. The number of statements increased (from two to seven) as well as their complexity. A statement of the lowest complexity would consist of one single sentence (e.g., A fridge is cold); a statement of a higher complexity would consist of two combined sentences (e.g., A fridge is cold *and* a kettle cools water); a statement with the highest complexity contained three combined sentences (e.g., A fridge is cold *and* a kettle cools water *and* a library has books).

The Afrikaans and Tswana versions consisted of meaningful true/false statements, reflecting familiar information for the Afrikaans and Tswana children respectively, such as "An alarm can make noise" for the Afrikaans test version and "A soccer team has 11 players" for the Tswana test version (soccer is the most popular sport among the Tswana children whereas rugby is most popular among the Afrikaans).

Figural fluid reasoning test. This individually administered test is based on the subtest *Situations* of the Snijders-Oomen Non-Verbal Intelligence Test, Revised (Snijders, Tellegen, & Laros, 1989) and relies on the same principle as Raven's Standard Progressive Matrices (Raven, Raven, & Court, 1998b). The child was shown a drawing with a missing part and had to choose from various answer options which piece best completed the drawing. The Afrikaans and Tswana versions consisted of 18 items and contained drawings of situations (at home, in school, on the streets) that were highly familiar to the Afrikaans and Tswana children, respectively. For example, for the Afrikaans test version, a drawing of a swimming pool was included, and the Tswana test version contained a drawing of a specific cooking procedure (two women preparing porridge in a three-legged pot). The first seven items had one missing part, the next seven items had two, and the last four items had three missing parts. Raven's Standard Progressive Matrices (Raven et al., 1998b) requires the child to complete a (meaningless) figural pattern and was collectively administered. The test was used as a reference point that did not reflect the Afrikaans or Tswana test version. Parts A, B, and C (36 items in total) were administered; the first two items of part A were used as example items.

Verbal fluid reasoning test. This collectively administered test consisted of 19 items. The child had to choose one word that did not go together with the other three (for the first 16 items) or the other two (for the last 3 items). Both the

Afrikaans (e.g., rugby – swimming – cricket – tennis) and Tswana version (e.g., grass – fire – three-legged pot – wooden spoon) contained items that were presumed to be highly familiar for the Afrikaans and Tswana children, respectively. The test examiner read the words out loud and the children could read along and circle their answer on an answer sheet.

Design

The children of each of the three groups (Afrikaans, urban Tswana, and rural Tswana) were divided into two subgroups; one for each of the two test versions (i.e., Afrikaans and Tswana), reflecting a 3x2 between subjects design. As far as possible, the subgroups were matched for sex, grade, and general level of school performance as estimated by the teachers. Raven's Standard Progressive Matrices was administered to all children to check the comparability of the subgroups that were selected for each of the test versions. An ANOVA with test version as independent variable and the score on the Raven as dependent variable showed that there were no significant differences, $F(1, 499) = 0.53$, $p = .47$. When looking at the performance differences on the Raven within each of the three groups, we found that both for Afrikaans and rural Tswana children, performance on the Raven did not significantly differ for the two different test versions. For the urban Tswana children however, the children selected for the Afrikaans test version performed significantly higher than those selected for the Tswana test version, $F(1, 179) = 5.41$, $p < .05$, partial $\eta^2 = .03$.

Familiarity questions

Content familiarity was assessed as a manipulation check of the perceived familiarity of both test versions. After each cognitive test administration, the children answered two content familiarity questions: 1) Were there any words/drawings that you did not know well in the task? (reverse keyed: *none, a few, many*); 2) How well did you know the words/drawings that we used in the task? (*not at all, a bit, very well*). For each of the two test versions (i.e., Afrikaans and Tswana), a factor analysis was performed on these two items for all tests of that particular test version (explained variances were 34%, 31%, and Cronbach's alpha values were .76 and .72, respectively). The factor scores were used as indicator of perceived content familiarity.

Socioeconomic status

Children were asked six questions as an indication of socioeconomic status (SES): 1) "Do you have your own room?" (*yes, no*); 2) "How many televisions are there in your house?"; 3) "Is there a microwave in your house?" (*yes, no*); 4) How many (cell)phones does your family (i.e., the people the child lives with) have?"; 5) "How many cars does your family have?"; 6) Do you have (reading) books at home?" (*yes, no*). One factor was extracted from these items (explained variance = 41%, Cronbach's alpha = .63) and the factor scores were used in further analyses.

Procedure

Eleven Afrikaans-speaking and eleven Setswana-speaking females were trained to administer the test battery. Seventeen were Psychology students, one had obtained her degree in Social Work, and four had completed high school. Consent for participation of the children in the study was obtained through the school principals. Individual testing took place in rooms that were made available by the schools and took about one hour for each child. One test examiner tested four children on average in a school day. Two tests (Raven's Standard Progressive Matrices and verbal fluid reasoning) were administered collectively in the classroom in about one hour; administration took place after all children had undergone individual testing.

Results

Results are described in three sections. We first present preliminary analyses on item bias, score standardization, background variables, reliability, and the manipulation check of perceived familiarity. This is followed by a validation of the cognitive structure that is hypothesized to underlie the test battery. Finally, a

Multivariate Analysis of Variance (MANOVA) is presented that tests for the effects of group and test version on the cognitive test scores.

Preliminary analyses

Item bias. Item bias (differential item functioning) was computed in a logistic regression procedure in which item scores were predicted on the basis of group membership (dummy coded), score level, and their interaction. The analyses showed that some items were biased; however, the effect sizes were small. We did not exclude any items from further analyses.

Score standardization. For each of the five cognitive tests, sum scores were computed for each of the two versions (combining Afrikaans, urban Tswana, and rural Tswana children). Analyses of these raw sum scores showed that significantly higher scores were obtained for the Tswana version of the attention test and the working memory test than for the Afrikaans version. To correct for these differences in difficulty levels, scores were standardized for each test version for all tests, thereby enabling a direct comparison of scores across versions.

Sex, grade, and socioeconomic status. In a MANOVA with sex as independent variable and the sum score of each test (standardized for test version) as dependent variables, we found a significant main effect of sex on short-term memory ($F[1, 499] = 4.56$, $p < .05$, partial $\eta^2 = .01$) and verbal fluid reasoning ($F[1, 499] = 4.80$ $p < .05$, partial $\eta^2 = .01$). Girls scored higher on these tests than boys. A MANOVA with grade as independent variable showed significant main effects for all test scores ($p < .01$, partial η^2 varying from .02 for verbal fluid reasoning to .04 for working memory). The socioeconomic level of the Afrikaans children was significantly higher than that of both Setswana-speaking groups, $F(2, 498) = 244.04$, $p < .01$, partial $\eta^2 = .50$. A MANOVA with SES as independent variable showed a main effect for all test scores ($p < .01$, partial η^2 varying from .03 for attention to .12 for verbal fluid reasoning), except for short-term memory. Higher SES was associated with higher cognitive test scores.

Reliability of cognitive tests. Table 5.1 gives an overview of the internal consistencies for all tests, specified for group and test version. Most values were acceptable to high. The highest values were found for the attention test; lowest values were found for the verbal fluid reasoning test.

Table 5.1

Internal consistencies of all cognitive tests for each group

	Group			
Test and version	Afrikaans	urban Tswana	rural Tswana	overall group
Short-term memory[a]				
Afrikaans	.60	.79	.78	.75
Tswana	.62	.81	.67	.74
Attention[a]				
Afrikaans	.95	.94	.93	.94
Tswana	.95	.93	.91	.93
Working memory[a]				
Afrikaans	.79	.78	.57	.78
Tswana	.75	.77	.71	.75
Figural fluid reasoning[b]				
Afrikaans	.58	.89	.69	.89
Tswana	.60	.81	.80	.77
Verbal fluid reasoning[b]				
Afrikaans	.38	.46	.58	.61
Tswana	.63	.71	.66	.70

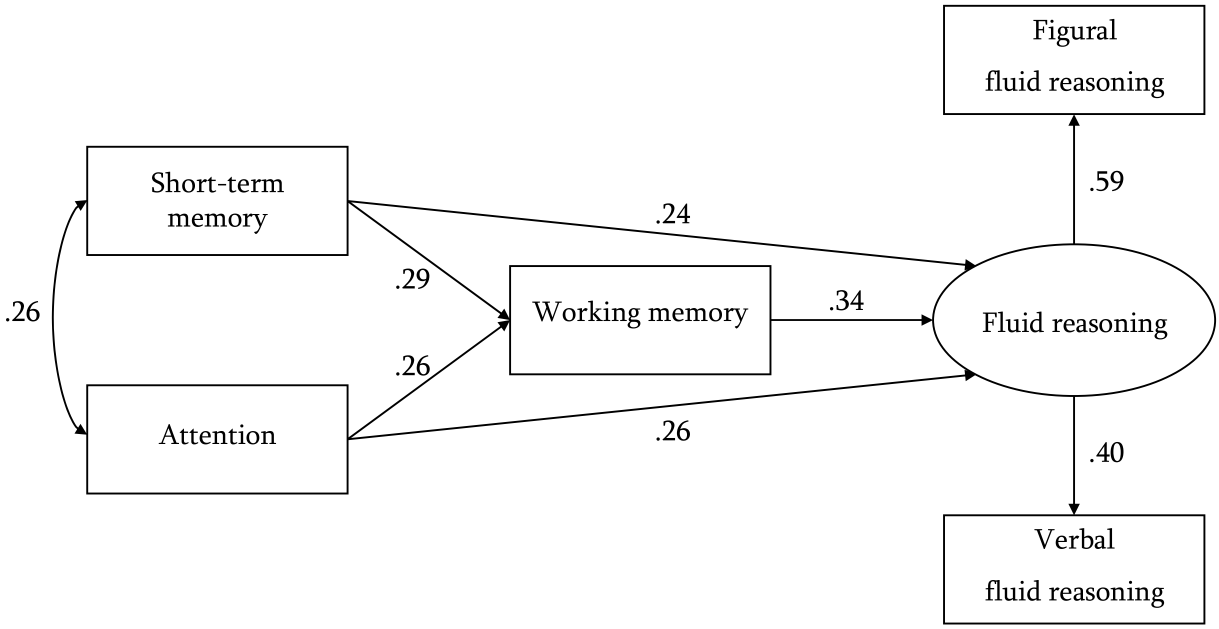
[a]Values are split-half reliabilities corrected for test length with the Spearman-Brown formula.

[b]Values are Cronbach's alpha values.

Manipulation check. Content familiarity was assessed as a manipulation check to establish whether the perceived familiarity was higher for the version of the own group than for the other version and to establish whether the perceived familiarity of a group's own test version was higher than the other groups' perceived familiarity of this same version. For each group, an ANOVA was performed with test version (two levels: Afrikaans and Tswana) as independent variable and the score on the familiarity questions (factor score) as dependent variable. For the Afrikaans group, the perceived familiarity of the Afrikaans version was significantly higher than that of the Tswana version $F(1, 159) = 35.89$, $p < .01$, partial $\eta^2 = .18$. For the urban Tswana group, there were no significant differences in perceived familiarity between the test versions, $F(1, 179) = 1.05$, $p = .31$, partial $\eta^2 = .01$. The rural Tswana group perceived the Tswana test version as more familiar than the Afrikaans version, $F(1, 157) = 9.71$, $p < .01$, partial $\eta^2 = .06$.

Subsequently, for each test version, an ANOVA was performed with group (three levels: Afrikaans, urban Tswana, and rural Tswana) as independent variable and the score on the familiarity questions (factor score) as dependent variable. Familiarity scores of the Afrikaans group were significantly higher than those of both Setswana-speaking groups on the Afrikaans test version, $F(2, 250) = 97.03$, $p < .01$, partial $\eta^2 = .44$ (post hoc analyses with Bonferroni criterion). On the Tswana test version, Afrikaans children scored significantly higher than the urban Tswana group, $F(2, 245) = 9.00$, $p < .01$, partial $\eta^2 = .07$. There were no significant differences between the Afrikaans and rural Tswana group and between the urban Tswana and rural Tswana group. Even though Afrikaans children reported a relatively high familiarity on the Tswana test version, the score differences between the Afrikaans group and the two other groups were smaller on the Tswana version than on the Afrikaans version, and the percentage of explained variance by familiarity was also substantially smaller. The manipulation check largely supported the adequacy of familiarity differences of the test versions.

A MANOVA with the score on the familiarity questions (factor score) as independent variable showed a significant effect for all cognitive test scores ($p < .01$, partial η^2 varying from .03 for short-term memory to .16 for figural fluid reasoning); as could be expected, higher familiarity was associated with higher test scores.



Note. All depicted parameter estimates are significant, $p < .01$

Figure 5.2
Cognitive structure underlying the data

Validity of cognitive structure

Using structural equation modeling (Arbuckle, 2008), the validity of the hypothesized cognitive structure (Figure 5.1) was tested. For the overall sample (N = 501), with scores standardized for test version and for group, we found an excellent fit ($\chi^2[2, N = 501] = 1.89$, $p = .39$, $\chi^2/df = .95$, GFI = 1.00, AGFI = .99, TLI = 1.00, CFI = 1.00, RMSEA = .00), confirming hypothesis 1a. The model is displayed in Figure 5.2. For both short-term memory and attention, the relation with fluid reasoning is partially mediated by the association between working memory and fluid reasoning. Multigroup analyses testing the invariance of the model across the test versions showed a good fit when all parameters were identical ($\chi^2[17, N = 501] = 28.76$, $p < .05$, $\chi^2/df = 1.69$, GFI = .98, AGFI = .96, TLI = .97, CFI = .97, RMSEA = .04), confirming hypothesis 1b. Multigroup analyses testing the invariance of the model across the three groups showed that only the unconstrained model (configural invariance) provided an excellent fit ($\chi^2[6, N = 501] = 7.16$, $p = .31$, $\chi^2/df = 1.19$, GFI = .99, AGFI = .96, TLI = .99, CFI = 1.00, RMSEA = .02), disconfirming hypothesis 1c. The reason for the lack of fit of models with more invariance constraints is not clear.

Table 5.2

MANOVA on cognitive test scores

Source	F^a	p	partial η^2
Group	Between subjects		
Short-term memory	8.34	.00	.03
Attention	2.38	.09	.01
Working memory	3.74	.03	.02
Figural fluid reasoning	13.62	.00	.05
Verbal fluid reasoning	11.33	.00	.04
Group x Test version			
Short-term memory	7.80	.00	.03
Attention	0.57	.57	.00
Working memory	10.07	.00	.04
Figural fluid reasoning	23.10	.00	.09
Verbal fluid reasoning	2.43	.09	.01
Error			
Short-term memory	(0.90)		
Attention	(0.95)		
Working memory	(0.86)		
Figural fluid reasoning	(0.77)		
Verbal fluid reasoning	(0.81)		

Note. The values for Test version are not displayed in this table because scores were standardized for each test version, leaving no significant main effects. Values in brackets represent mean square errors.

[a] $df(2, 495)$

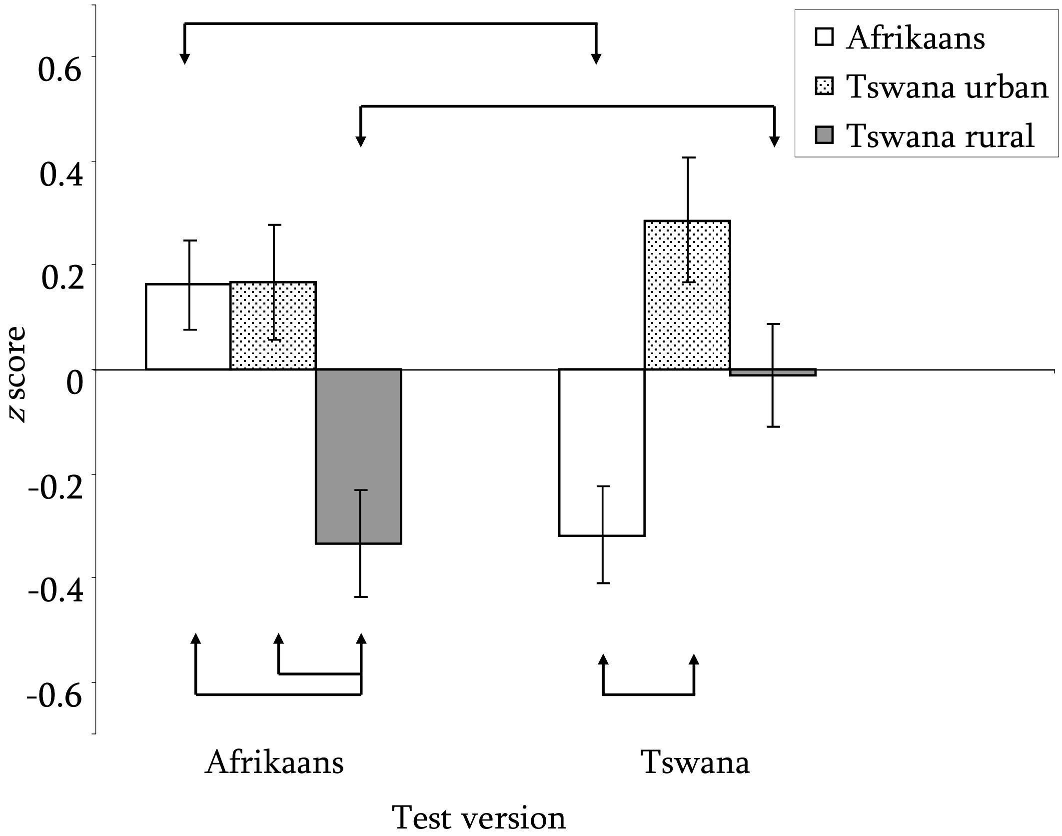
MANOVA on cognitive test scores

To correct the cognitive test scores for the effects of sex, grade, and SES, we first performed a MANOVA with these variables as independent variables and the scores on each test (standardized for test version) as dependent variables, and saved the residual scores. Then a MANOVA was performed with test version (two levels: Afrikaans and Tswana) and group (three levels: Afrikaans, urban Tswana, and rural Tswana) as independent variables, and the residual scores as dependent variables (see Table 5.2). Test version had no significant effect, due to the standardization of the scores for each test version. Group showed a significant effect on all test scores except for attention. Interactions between group and test version were significant for short-term memory, working memory, and figural fluid reasoning. They were not significant for attention and verbal fluid reasoning.

Figures 5.3 to 5.7 show the mean *z* scores and the significance of the score differences between the groups and between the test versions for each cognitive test in univariate tests. Score differences were expected to increase with cognitive complexity. When comparing Figures 5.3 to 5.7, it becomes clear that hardly any score differences were found for attention and larger differences were found for working memory and figural fluid reasoning. Contrary to our expectations, quite large differences were found for short-term memory and small differences for verbal fluid reasoning. Hypothesis 2 was only partially confirmed.

Table 5.3 provides an overview of the (dis)confirmation of hypotheses 3 to 9. Afrikaans children performed significantly better on their own test version than on the Tswana test version for short-term memory, working memory, and figural fluid reasoning, and performed equally on both versions for attention and verbal fluid reasoning. This confirms hypothesis 3 for all tests except short-term memory (we expected small or no differences) and verbal fluid reasoning (we expected large differences). The rural Tswana children scored significantly higher on the Tswana test version than on the Afrikaans version for short-term memory, working memory, and figural fluid reasoning. For attention and verbal fluid reasoning, scores did not significantly differ for the Afrikaans and Tswana version. These findings confirmed hypothesis 4 for all tests with the same two exceptions: short-term memory (we expected small or no differences) and verbal fluid reasoning (we expected large differences). The urban Tswana children performed equally well on the Afrikaans and Tswana test version for all tests, confirming hypothesis 5 for all tests.

Afrikaans children scored significantly higher than the Tswana children on the Afrikaans test version for each of the five tests, except for the attention test that did not show significant differences. Hypothesis 6 was confirmed for all tests except for short-term memory (we expected small or no differences); yet, the difference was only significant in comparison with the rural Tswana group. This group did not score significantly higher than the Afrikaans children on any of the tests of the Tswana test version (disconfirming hypothesis 7 for the working memory and fluid reasoning tests, for which we expected large differences). Urban Tswana children scored in between Afrikaans and rural Tswana children on the Afrikaans test version of all tests, except for short-term memory; hypothesis 8 is confirmed for all tests except for short-term memory. Urban Tswana children did not score in between Afrikaans and rural Tswana children on the Tswana test version of short-term memory and verbal fluid reasoning, disconfirming hypothesis 9 for these tests. For the other tests of the Tswana version, there were no significant differences between the groups (confirming hypothesis 9 for these tests).



Note. Arrows indicate significant differences (post hoc analyses with Bonferroni criterion). Significant differences between test versions: $F(1, 159) = 13.95$, $p < .01$ (Afrikaans group) and $F(1, 157) = 5.18$, $p < .05$ (rural Tswana group). Significant differences between groups: $F(2, 250) = 8.07$, $p < .01$ (Afrikaans version) and $F(2, 245) = 8.23$, $p < .01$ (Tswana version).

Figure 5.3

Mean z scores for groups and test versions for short-term memory

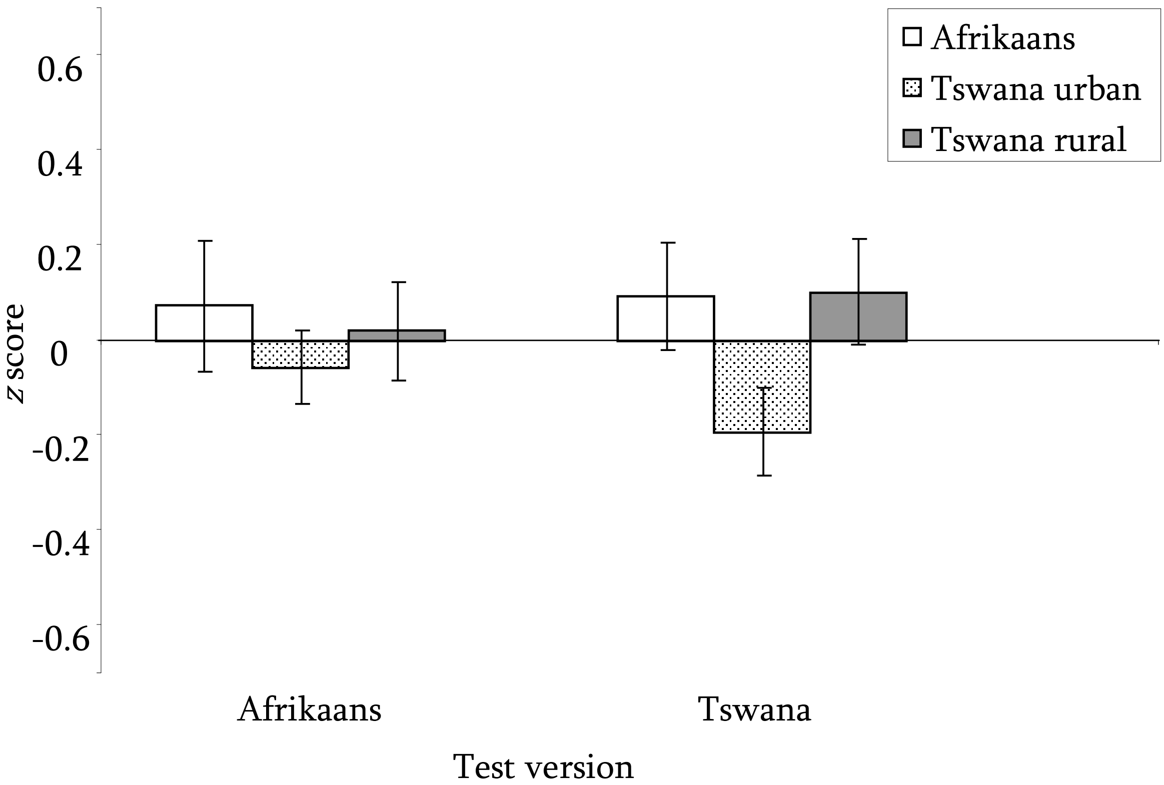
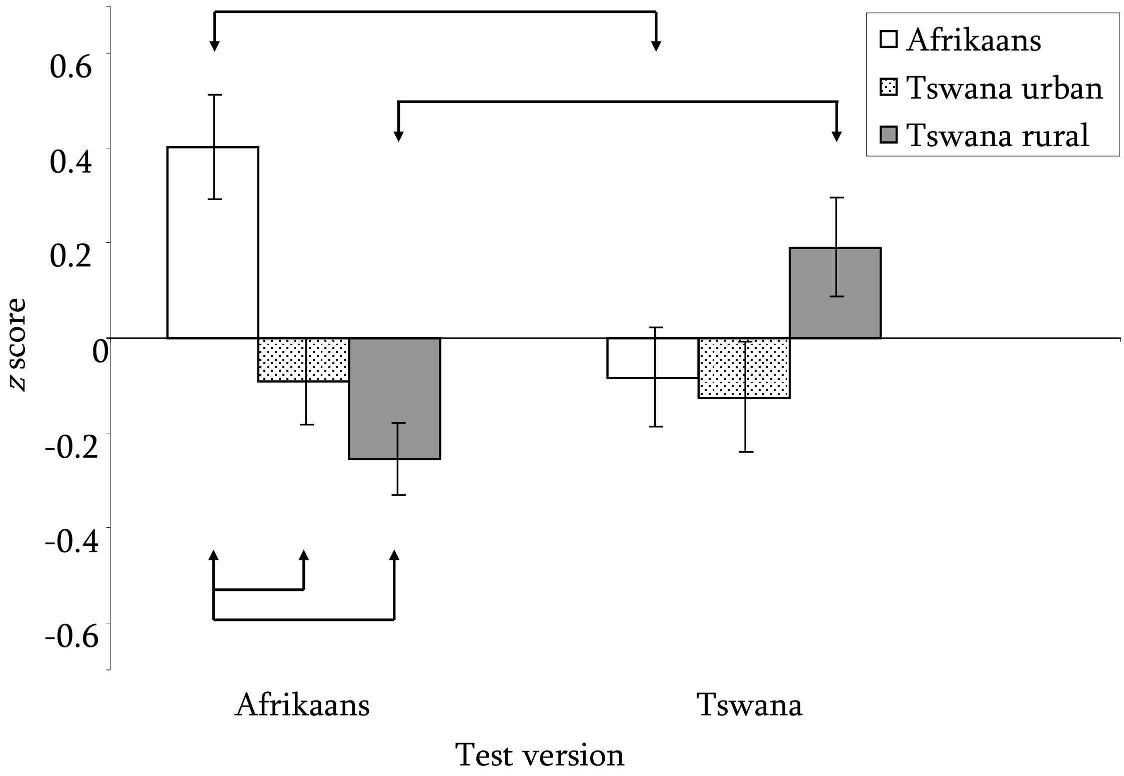


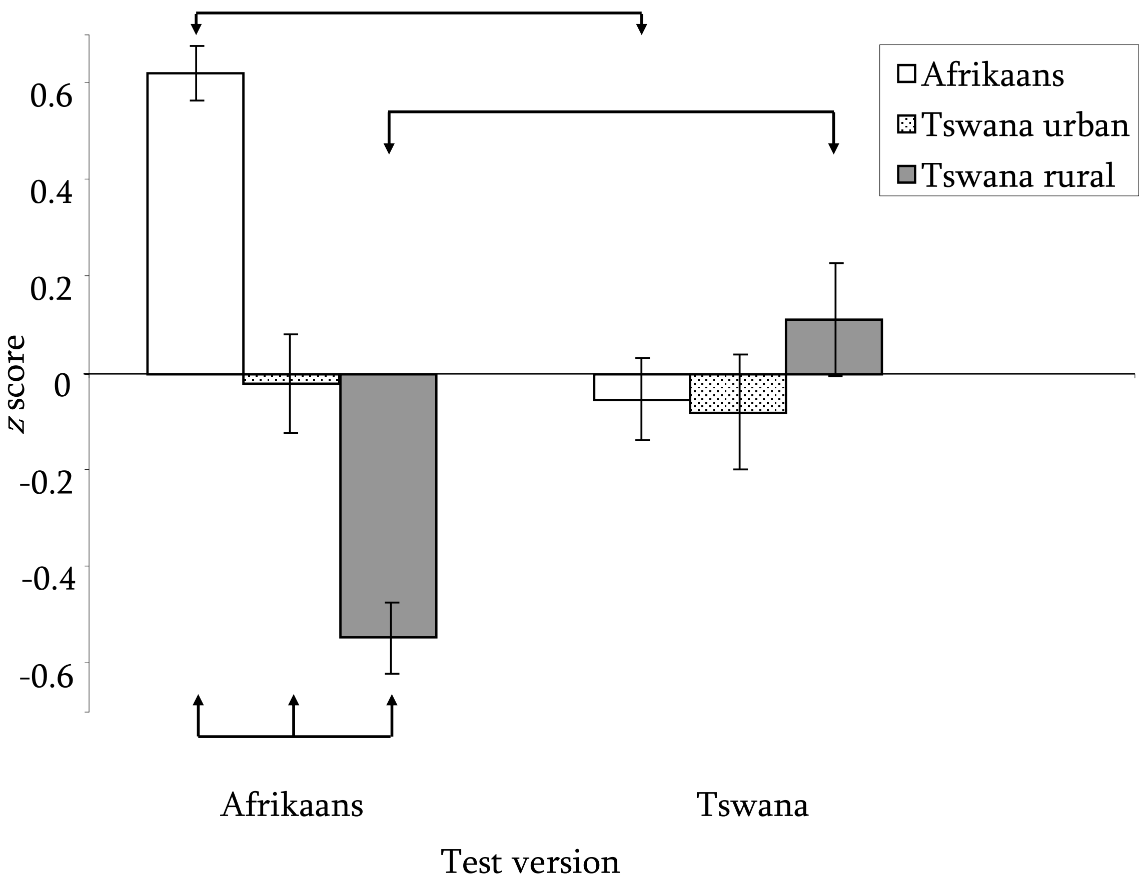
Figure 5.4
Mean z scores for groups and test versions for attention



Note. Arrows indicate significant differences (post hoc analyses with Bonferroni criterion). Significant differences between test versions: $F(1, 159) = 10.22$, $p < .01$ (Afrikaans group) and $F(1, 157) = 12.15$, $p < .01$ (rural Tswana group). Significant differences between groups: $F(2, 250) = 12.82$, $p < .01$ (Afrikaans version).

Figure 5.5

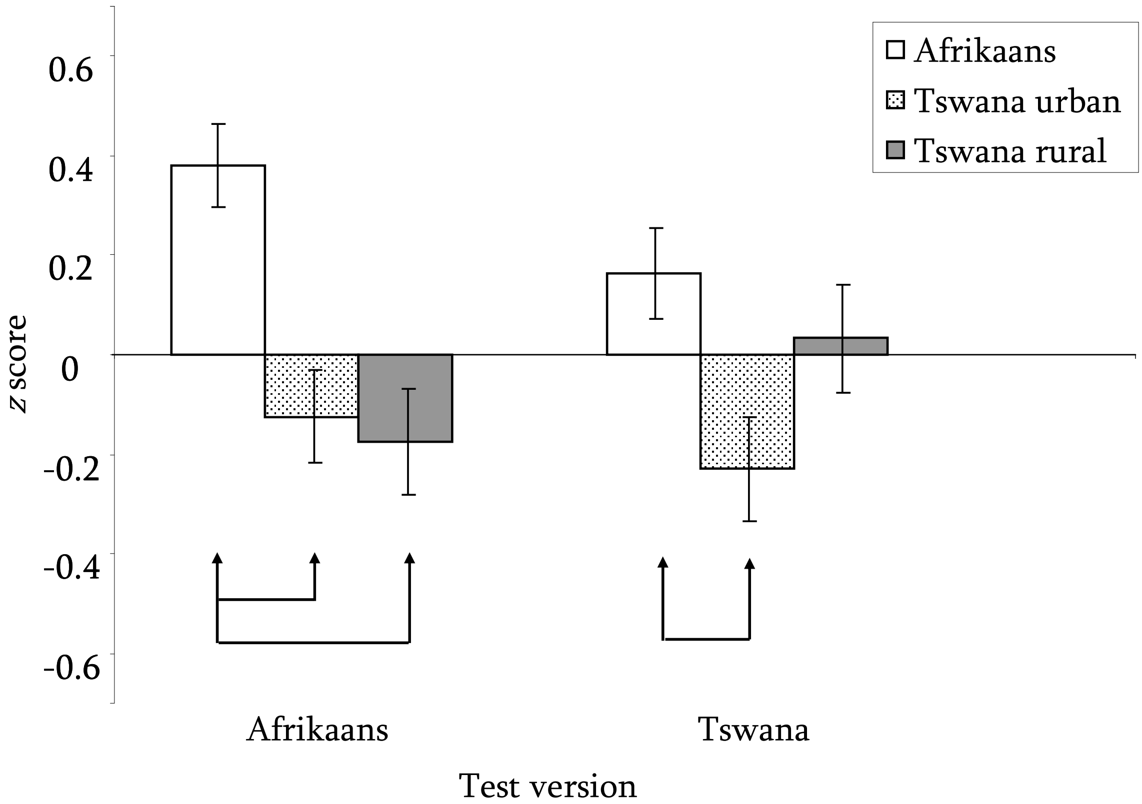
Mean z scores for groups and test versions for working memory



Note. Arrows indicate significant differences (post hoc analyses with Bonferroni criterion). Significant differences between test versions: $F(1, 159) = 43.57$, $p < .01$ (Afrikaans group) and $F(1, 157) = 23.47$, $p < .01$ (rural Tswana group). Significant differences between groups: $F(2, 250) = 47.95$, $p < .01$ (Afrikaans version).

Figure 5.6

Mean z scores for groups and test versions for figural fluid reasoning



Note. Arrows indicate significant differences (post hoc analyses with Bonferroni criterion). Significant differences between groups: $F(2, 250) = 10.25$, $p < .01$ (Afrikaans version) and $F(2, 245) = 3.96$, $p < .05$ (Tswana version).

Figure 5.7

Mean z scores for groups and test versions for verbal fluid reasoning

Table 5.3

Confirmation of hypotheses 3 to 9 for all cognitive tests

	Expected score pattern		Cognitive test				
	Group	Test version	Short-term memory	Attention	Working memory	Figural fluid reasoning	Verbal fluid reasoning
3	Afrikaans	Afrikaans ≥ Tswana	no	yes	yes	yes	no[a]
4	Rural	Afrikaans ≤ Tswana	no	yes	yes	yes	no[a]
5	Urban	Afrikaans = Tswana	yes	yes	yes	yes	yes
	Test version	Group					
6	Afrikaans	Afrikaans ≥ urban & rural Tswana	no[b]	yes	yes	yes	yes
7	Tswana	Afrikaans ≤ rural Tswana	yes	yes	no[a]	no[a]	no[a]
8	Afrikaans	Afrikaans ≥ urban Tswana ≥ rural Tswana	no	yes	yes	yes	yes
9	Tswana	Afrikaans ≤ urban Tswana ≤ rural Tswana	no	yes	yes	yes	no

Note. Hypotheses 3, 4, 6, and 7 are correct for short-term memory and attention when score differences are absent or small, for working memory when they are large(r) and for fluid reasoning when they are largest. [a]Score differences were not significant. [b]The difference was significant for the Afrikaans and rural Tswana group.

CCH predicts that the cross-cultural score differences are explained by the cultural complexity of the tests (conceptualized in the present study as the extent to which test content is more familiar to one of the compared groups). In analysis of variance terms, CCH predicts that only *disordinal* interactions are found between test version and group for the tests with high cognitive complexity: it depends on the test version which group scores highest. SH predicts that cross-cultural score differences can be explained by differences in general cognitive ability. This implies that, according to SH, only main effects for group and, possibly, *ordinal* interactions between test version and group are expected for the high cognitively complex tests: one group scores consistently higher than the other(s). In line with expectations, the largest score differences were found for the (theoretically) more cognitively complex tasks. Most significant interactions on these tests were found for the Afrikaans and rural Tswana group and were disordinal, in line with CCH and not with SH. Differences between the Afrikaans and rural Tswana group were largest for the Afrikaans version of figural fluid reasoning ($F(1, 160) = 155.77$, $p < .01$, partial $\eta^2 = .49$, Cohen's $d = 1.96$). Figural fluid reasoning showed the largest differences between test versions in both the Afrikaans group ($F(1, 159) = 43.57$, $p < .01$, partial $\eta^2 = .22$, $d = 1.04$) and the rural Tswana group $F(1, 157) = 23.47$, $p < .01$, partial $\eta^2 = .13$, $d = -0.76$).

Discussion

Where do cross-cultural differences in cognitive test scores come from? Spearman's Hypothesis (SH) holds that the differences are mainly caused by cross-cultural differences in cognitive abilities; however, we expect them to be dependent on the cultural rather than cognitive complexity of a test (Cultural Complexity Hypothesis, CCH). In the current study the content familiarity of five cognitive tests was manipulated to examine its effect on test performance. Two test versions were created, an Afrikaans and a Tswana version. The tests were administered to groups of (urban) Afrikaans children, (urban) Tswana children from the same area as the Afrikaans children, and (rural) Tswana children from an area that is relatively isolated from Afrikaans culture. We found an excellent fit of our hypothesized cognitive structure when analyzing the sample as a whole. The relation of both short-term memory and attention with fluid reasoning was partially mediated by working memory. Only configural invariance could be established in a comparison of the factor structure for the three groups (Afrikaans, urban Tswana, and rural Tswana). Afrikaans and rural Tswana children generally performed better on the test version that was designed for their own group than on the other test version. The urban Tswana group did not score differently on the Afrikaans and Tswana test version, showing that these children have enough knowledge of both cultures to perform equally on both versions. Afrikaans children generally scored higher on the Afrikaans version than the Tswana children. Tswana children however, did not significantly score higher than the Afrikaans children on the Tswana test version. Nevertheless, most performance differences between the groups were smaller on this version than on the Afrikaans version; the rural Tswana group scored significantly higher on the Tswana version than on the Afrikaans version and the Afrikaans group scored significantly lower. We can conclude that our results support the idea that the content familiarity of tests was an important moderator of cross-cultural differences in test scores in that children generally performed better on the test version that was designed for their own group than on another test version.

The short-term memory test appeared to be more sensitive for group differences than expected, given the test's low cognitive complexity. Urban Tswana children scored highest on both versions of this test. We could not capture any educational characteristics that could explain these findings (such as specific training of memory abilities or a stronger reliance on rote learning in the urban Tswana group as compared to the other groups). Overall, the largest score differences were found between the Afrikaans and the rural Tswana groups for the working memory and figural fluid reasoning tests. These tests are seen as cognitively more complex than the attention and short-term memory tests. SH would predict that on two test versions with a comparable level of cognitive complexity, regardless of content, score differences between groups are in the same direction (i.e., interactions between test version and group are ordinal). However, on inspection of our results, there were significant score differences on the Afrikaans version and no significant differences on the Tswana version (interactions were disordinal). One could argue that the difficulty level of the Tswana version was lower than that of the Afrikaans version; however, this was only the case for the working memory test. More importantly, regardless of the level of difficulty, the Afrikaans children performed significantly lower on the Tswana version than on the Afrikaans version. Therefore, we conclude that our findings are not in line with SH, and that it is cultural rather than cognitive complexity that explains most differences between groups, providing support for CCH.

Our study fits in a pattern of studies that have given arguments to question the validity of SH. The first type of argument focuses on the statistical analyses applied to test SH that are said to be too lenient (see Dolan et al., 2004). The second type of argument concerns the confounding of cognitive complexity with cultural complexity in current tests of SH. A high loading on a general cognitive ability factor does not merely imply a high cognitive complexity, but usually goes together with higher cultural complexity. Confirmations of SH that have been reported in the literature (e.g., Hartmann et al., 2007; Lynn & Owen, 1994; Te Nijenhuis & Van der Flier, 1997) may be based on this confounding in the data. We confirmed findings by Helms-Lorenz et al. (2003) which indicated that SH can only be tested when cultural complexity and cognitive complexity are both varied independently. Data from the present study and from Helms-Lorenz et al. show that when these types of complexity are unconfounded, SH is not supported.

In addition to experimentally manipulating the content familiarity of the tests by creating two versions, familiarity questions were used to check the perceived familiarity of both versions. The content familiarity questions served their purpose of a manipulation check relatively well; yet their validity could be challenged. First, social desirability could have played a role in that children indicated to know certain words or drawings because they believed they were expected to know these. Second, children may not have good insight in their familiarity with stimuli as compared to tasks. Some children found it difficult to independently evaluate the complexity of stimuli (words and drawings) and of the task (what had to be done with the stimuli). Content familiarity appeared very difficult to measure. Rather than merely relying on self-report, it would be an idea to include a more objective measure of content familiarity. A test exposing children to various types of test content and measuring their reaction time in manipulating this content might circumvent the validity issues.

This study has two limitations. First, the results show that the Afrikaans and Tswana versions of the verbal fluid reasoning test were not culturally loaded to the extent that they could show differences between the groups. It was difficult to construct items that tap cultural complexity to the same degree and show substantial variation in difficulty; this lack of coherence could have resulted in the low internal consistencies. Second, for each of the four cognitive abilities reflected in our test battery, only one test was used (except for fluid reasoning, for which two tests were used). To find more unequivocal support for the cognitive structure underlying the tests (as displayed in Figure 5.1), probably more tests would need to be included.

Our study has some practical implications. Unfamiliar test content can have a significant negative effect on a child's test performance, possibly providing an inaccurate estimation (underestimation) of the child's ability. Therefore, tests need to be selected on the appropriateness of their content whenever possible. The content familiarity of a cognitive test should be taken into account when evaluating a child's performance or when cross-culturally comparing scores. Dynamic testing procedures provide an opportunity to overcome at least some cross-cultural differences in familiarity. The more traditional static tests evaluate the child's current cognitive performance, which is influenced by the test's cultural complexity and might not be a good reflection of the underlying cognitive ability. However, dynamic testing consists of a pre-test phase, a training phase, and a post-test phase (e.g., Sternberg & Grigorenko, 2002) and is focused on the child's ability to learn, reflected in the score difference between the pre- and post-test. The child's cognitive ability as measured by dynamic testing is less influenced by cultural complexity (Hessels, 1997; Sternberg et al., 2002). The combination of appropriate item content and dynamic testing is promising in closing the cross-cultural gap in test scores.

The present study also has a theoretical implication. Cognitive abilities are domain dependent (i.e., their expression is dependent on aspects such as the type of cognitive task and the familiarity of its stimuli), notably the more complex abilities; however, SH does not consider domain features as relevant in the explanation of cross-cultural differences in cognitive test scores. Neo-Piagetian psychology (e.g., Demetriou, Shayer, & Efklides, 1992) and cognitive psychology (e.g., Keane & Eysenck, 2005) include domain features in their models. Cognitive models that accommodate cross-cultural differences in abilities should also incorporate these features (such as stimulus familiarity). Models of cross-cultural differences in cognitive functioning can only be comprehensive when they address the domain dependence of these differences.

Chapter 6

Summary and General Discussion

The broad underlying question I address in this dissertation is how culture and children's cognition are related. Taking into account bias in assessment procedures, I have shed light on this relation from three angles. The first focused on detecting and reducing bias in cognitive tests (Chapters 2 and 3); the second examined cultural influences on the relation between a child's home environment and cognition (Chapter 4); the third manipulated bias by varying the content familiarity of tests to answer the question of why some cognitive tests show more cross-cultural differences than others (Chapter 5).

Summary of chapters

Chapter 2 described and applied a judgmental (qualitative) procedure for cognitive test adaptations. The procedure consisted of iterations of translating, piloting, and modifying the instrument. Five types of adaptations for cognitive instruments were distinguished, based on the underlying source: construct, language, culture, theory, and familiarity, respectively. The proposed procedure was applied in an adaptation of 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 were needed of test instructions, item content of both verbal and non-verbal tests, and item order. It was concluded that the qualitative approach adopted here could adequately identify various problems with the application of the KABC-II in this sample that would have remained unnoticed with a close translation of the original instrument.

Chapter 3 evaluated the psychometric adequacy of the extensive adaptation of the KABC-II. The subtests showed high reliabilities in a sample of 598 children, the Cattell-Horn-Carroll model underlying the original KABC-II was largely replicated, and external relations with demographic characteristics and an achievement measure were consistent with expectations. The subtests showed relatively high loadings on the general cognitive factor, presumably due to the high task novelty and hence, cognitive complexity of the tests for the children. The findings support the suitability and validity of the KABC-II adaptation. It was concluded that test adaptations can only be adequate if they meet both judgmental (qualitative) and statistical (quantitative) adaptation criteria.

Chapter 4 examined the associations between the home environment and cognitive performance of the Indian children. The primary caregivers of 532 children that underwent the KABC-II test administration (see Chapter 3) were interviewed. The path model supported the applicability of environment-cognition relations as suggested by the investment model and family stress model, both mainly based on Western data. More proximal variables (caregiver's mental health, social support, family conflict, and parental behaviors) showed stronger associations with cognitive performance than more distal variables (socioeconomic status); the link between the latter and cognitive outcome was fully mediated by the more proximal variables. Only the direct involvement of parents with their child through providing stimulation and structure was directly related to outcome. Interventions to improve developmental outcomes should therefore target the most proximal parental behaviors.

In Chapter 5 cross-cultural differences in cognitive test scores were hypothesized to depend on a test's cultural complexity (Cultural Complexity Hypothesis: CCH) as reflected in its content familiarity rather than on its cognitive complexity (Spearman's Hypothesis: SH). The content familiarity of tests assessing short-term memory, attention, working memory, and figural and verbal fluid reasoning, was manipulated by constructing test versions with an item content derived from either Afrikaans or Tswana culture in South Africa. Both test versions were administered to children of both cultures. The sample consisted of 161 urban Afrikaans-speaking, 181 urban and 159 rural Setswana-speaking children. The results disconfirmed SH and supported CCH; children generally performed best on the test version that was designed for their own group, particularly for the cognitively and culturally complex working memory and figural fluid reasoning tests. Content familiarity is an important moderator of cognitive test performance that should be taken into account, particularly in cross-cultural comparisons of scores.

Discussion of main findings

With Chapters 2 and 3 I have shown that cognitive instruments developed and validated in a Western context can be adapted for use in a completely different cultural context, even though the process is extensive and time-consuming. Many more test aspects than anticipated required adjustments; bias is everywhere. The adaptation did not only deal with cultural differences between the original (American) context and the target (Indian) context; linguistic and socioeconomic differences were also addressed. Most of the adaptations were driven by the familiarity of item content and of tasks, implying that familiarity is the main point of concern when adapting a test. To properly improve a test's suitability for the target context, various experts need to be included in the study, such as psychologists, linguists, teachers, and parents. Both qualitative and quantitative procedures are required to evaluate the success of any adaptation.

Chapter 4 indicated that the salience of certain environment-cognition relations can differ across cultures and across socioeconomic strata. In that sense, bias does not only affect the generalizability of a cognitive test, but also the generalizability of relations between environmental and outcome variables. A distinction can be made between factors in the home environment that are either distal or proximal to a child's cognitive performance. Interventions that focus on distal factors by improving a family's socioeconomic status (by educating caregivers or by an increase in family income) would only have effect on a long term. Interventions targeting the most proximal parental behaviors by stimulating parents' direct involvement with their children seem most effective and also most feasible on a short term.

Chapters 2 and 3 already stressed the importance of familiarity for the appropriateness of a test. Chapter 5 provided further support for this finding by showing the impact of content familiarity on a child's cognitive test performance. Strong indications were found for the role of cultural complexity in explaining cross-cultural score differences. Knowing that it is the content familiarity of a test rather than a cross-cultural difference in general cognitive ability that largely explains score differences, makes it even more important to have culturally appropriate instruments.

Implications

Adoption, assembly, or adaptation

The findings from Chapters 2 and 3 are in line with those from Abubakar et al. (2007) and Holding et al. (2004), who also demonstrated the necessity and utility of an adaptation of a Western instrument for a non-Western context. It is common, though, to directly translate tests into the target language without further adjustments (Van de Vijver & Poortinga, 2005). When tests are used in another context without making changes in any of the test aspects or procedures, the validity of the test scores is questionable in case the cultural gap between the culture in which the test originated and the target culture is large. This dissertation shows that adaptation can be a useful tool to bridge this gap. I strongly suggest that the choice between adoption, adaptation, and assembly, is an integral part of any test selection procedure and is justifiable.

Choosing to adapt. I chose for the option of adaptation to be able to retain test elements that are suitable for the target context and adjust the elements that are culturally inappropriate. Adoption would not have been an option since it is only appropriate when there are small (linguistic and cultural) differences between the original and target context. One might wonder why I preferred the option of adaptation over developing an appropriate test from scratch in the target culture itself (i.e., assembly), since in the latter case, the cultural suitability of tasks and response formats is largely guaranteed. An issue with locally developed tests, however, is that they are usually not based on a solid (i.e., cross-culturally tested) cognitive model. Without having a well-founded underlying structure, it is difficult to evaluate what the test actually measures; the validity is doubtful since there is no clear point of reference. There are also some downsides to the use of test adaptations. First, there is a need for specialized statistical procedures that can deal with nonoverlap of items across cultures such as item response theory and structural equation modeling (Van de Vijver & Leung, 1997). Second, adaptation does not necessarily take into account that certain cognitive abilities that are evaluated as important in the culture from which the test originates might be less relevant and less valued in the target culture. The universalistic approach that was adopted in this study does, however, not argue that all cognitive abilities have the same importance or the same manifestations across cultures; it merely emphasizes the universality of the underlying cognitive functions. The current study supported the validity of the chosen approach; presumably aided by the thorough test adaptations, the underlying CHC model was replicated.

Universalism and adaptation. From the discussion on the costs and benefits of the three methods of test transfer, I can conclude that the method of adaptation is particularly suitable for studies that are conducted from a universalistic point of view. Many studies are based on this point of view, leaving ample opportunity for successfully applying adaptations. Assembly would be more appropriate when researchers take a relativistic perspective. Adoption of tests is preferred over adaptation when the gap between original and target culture is negligibly small, regardless of the researcher's perspective of absolutism, universalism or relativism.

Explaining cross-cultural score differences

Chapter 5 focused on detecting the source of cross-cultural differences in cognitive test scores. Spearman's Hypothesis has been widely studied and has been widely confirmed. Criticism has also been expressed toward the role of cognitive complexity in explaining score differences. I have shown that cultural complexity played a more important part in explaining the differences between groups than the level of cognitive ability that is required to successfully complete a test. It is understandable that the explanations of cognitive and cultural complexity are often confounded. There is a correlation between a test's cognitive and cultural complexity; the higher the cognitive complexity of a test, the more cultural (contextual) information is usually needed to perform well.

The importance of cultural complexity in explaining cross-cultural differences does not imply that cognitive complexity does not matter. When one of the studied groups has very little experience with certain cognitive tasks and less training than the other groups in the cognitive abilities reflected in those tasks, performance of its members is negatively influenced. As as consequence, cross-cultural score differences are not merely a reflection of differences in familiarity with test content but also of differences in skills as a result of differences in cognitive ability training. This implies that research aimed at addressing cross-cultural score differences should take both explanations (cultural and cognitive complexity) into account and should be careful in drawing conclusions on the importance of one as compared to the other.

Interventions fostering child development

Chapters 2, 3, and 5 emphasized the importance of appropriate assessment instruments. Chapter 4 illustrated one of the purposes for which culturally appropriate tests are of major importance, namely for obtaining an accurate measure of factors in the child's home environment that affect developmental outcomes. The chapter showed that the aspects in the home environment that are most proximal to the child's cognitive performance are the ones that matter the most. Making mothers aware of the influence they can have on their children's cognitive outcome constitutes an important target for interventions that foster child development. Direct maternal involvement with the child through providing structure and stimulation appears to be essential. However, views on the importance of and the responsibility for stimulating a child's cognition presumably show cross-cultural differences. The Indian caregivers of low socioeconomic background as described in Chapter 4 do perhaps not see it as their duty to promote their child's cognitive development but defer this task to school teachers. Bronfenbrenner (1979) identified three microsystems (home, school, and peers) that affect a school going child's development. Combined interventions in all three areas might have a much larger impact on child development than an intervention in one of them. Nevertheless, this dissertation contributes to the knowledge on targets for intervention in one of the microsystems, by identifying crucial factors in the home environment of the low SES Indian child.

Integrating culture, cognition and bias

In the introduction I presented a simple model describing the relation between the three key concepts of this dissertation, being culture, cognition, and bias. Studies on the relation between the first two concepts can only be done by taking into account the third. The model is displayed once more in Figure 6.1. The studies described in Chapters 2 to 5 focused on the direct living environment of the child, as a reflection of a broader cultural context. Bias was examined, incorporated, or dealt with in the measurement of various parenting behaviors in the day to day life of the child as well as in the measurement of cognitive abilities. If there is no bias in any of the assessment instruments, measurements of the child's environment and cognition reflect the actual environment and actual cognitive abilities. This unbiased assessment is the ultimate aim of any research, especially in a cross-cultural context. Because a complete absence of bias is highly unlikely, I emphasize the necessity to take this concept into consideration in any study, regardless of the often high face validity of assessment instruments.

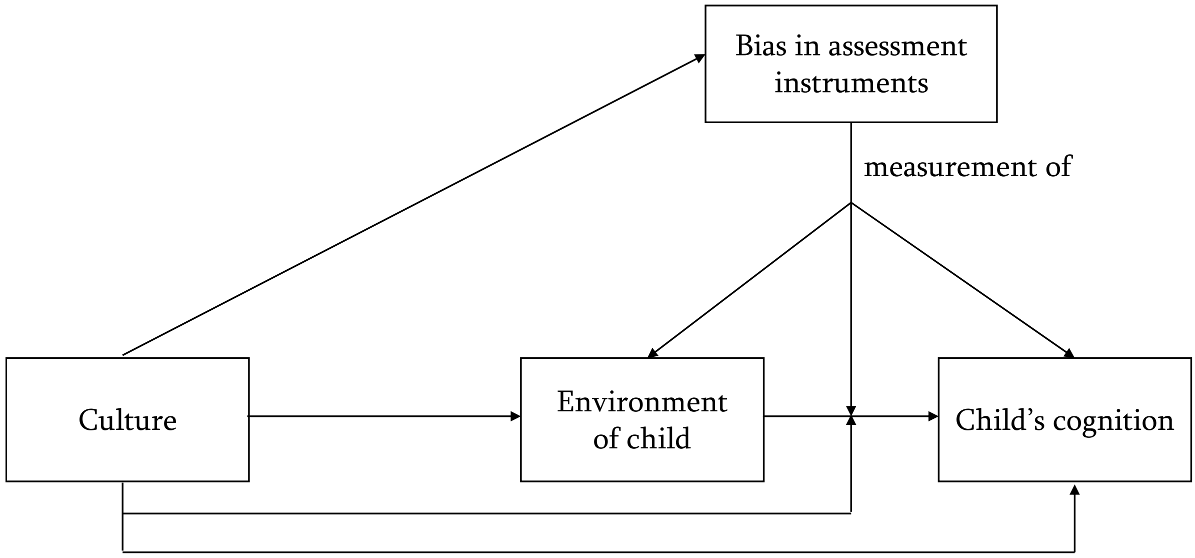


Figure 6.1
A model of culture, cognition, and bias

The future of cognitive testing

How bad is bias?

Whereas Chapters 2 and 3 focused on the reduction of bias and thereby viewed it as something negative that should be avoided or removed, Chapter 5 showed that bias is not necessarily bad. A culture free test does not exist; the closest alternative may seem a culture fair test, for which one group does not have better chances than another group to perform well. Decontextualizing a test (i.e. reducing bias) seems to lead to the development of such a culture fair test. However, even if groups are equally exposed to testing situations and certain types of tasks, they usually differ in their familiarity with the task content (e.g., drawings or words). It is impossible to obtain a test that is equally familiar to all included groups, implying that it is impossible to completely eliminate bias. Reducing bias against a cultural group by adapting a test usually implies introducing bias against other groups. Apparently, decontextualizing does not provide the desired result and we should rather focus on contextualizing (i.e., taking into account bias instead of trying to avoid bias) by ensuring that tests are as familiar as possible for the target groups. Because bias cannot be eliminated, contextualizing seems to be the best option.

Contextualized tests and cross-cultural research: do they go together?

Constructing a test in such a way that the content is contextualized seems hard to combine with cross-cultural research that focuses on comparing scores across groups that substantially differ from each other culturally. The downside of this approach after all is that not one and the same test is administered to different groups; the equivalence of these test versions then needs to be tested to draw conclusions on the comparability of scores. When test versions are being constructed for each of the target groups of a study, independent raters can be asked to rate the cultural complexity of the item content so as to obtain comparable levels for all test versions. In addition, all versions could be administered to a representative sample of all included groups. When the difficulty level of the test versions is found to be equal (i.e., the average scores for all groups combined are similar across test versions), these versions can be used in cross-cultural comparisons of these groups. In the actual data collection, the abilities of members of the included groups are then only assessed with the test version that was designed for their own group. The question, however, remains whether these test versions can really be treated as each other's (context-dependent) substitutes.

By including more tests of the same cognitive ability, a clearer and more objective view of this ability can be obtained. It would be desirable but it does not seem feasible to include more than one culturally adapted instrument of the same ability. A more feasible alternative would perhaps be to also include a decontextualized test (i.e., a test with the lowest possible level of bias) and administer this same test to all included groups. Correlations with the scores on this decontextualized test should not differ substantially between both contextualized versions. The best indication of cognitive ability might then be the average of the scores on the contextualized and the decontextualized test. Even though this process of obtaining indications of cognitive abilities is laborious, I do believe that this is the direction that should be taken when the aim is to assess cognitive abilities in a culturally sensitive way. Besides properly selecting the appropriate tests or making useful cultural adaptations, another point of attention is the use of dynamic as opposed to static cognitive tests.

Dynamic versus static testing

A traditional static test measures current performance, which is only a snap-shot of the underlying cognitive abilities. As I emphasized in this dissertation, current performance can be affected by many factors (captured by the term "bias") that distort the representation of actual abilities. Dynamic testing has been proposed to provide a more valid (i.e., less biased) indication of these actual abilities. The procedure commonly consists of a pre-test phase, a training phase, and a post-test phase (e.g., Sternberg & Grigorenko, 2002). The score difference between the pre- and post-test reflects the child's ability to learn and is the focus of this type of testing. Dynamic testing has been proven successful with children that are disadvantaged due to factors such as cultural differences, lack of educational opportunities, or learning disabilities (Hessels, 1997; Sternberg et al., 2002; Tzuriel, 2001). I emphasize, though, that even for dynamic testing procedures, the content of the test is relevant. Children that are familiar with the content might have an advantage over the ones that do not, even though the test focuses on learning ability; this learning ability in itself can be influenced by familiarity as well. Nevertheless, dynamic testing combined with appropriate item content can contribute substantially to closing the cross-cultural gap in cognitive test scores.

To conclude

How are culture and children's cognition related? There is no simple answer to the main question underlying this dissertation. They are related, that is for sure. This relation becomes visible when cross-culturally comparing cognitive test scores and when identifying targets for intervention in a child's home environment. Culture is everywhere, bias is everywhere. It depends on the frame of reference to what extent the influence of culture or the presence of bias is a problem. Nevertheless, I conclude that a (cognitive) test is at home in the country or context in which it was either originally developed or adapted for. There is no place like home.

References

Abubakar, A., Van de Vijver, F. J. R., Mithwani, S., Obiero, E., Lewa, N., Katana, K., et al. (2007). Assessing developmental outcomes in children from Kilifi, Kenya, following prophylaxis for seizures in cerebral malaria. *Journal of Health Psychology, 12*, 417-430.

Ackerman, P. l., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin, 131*, 30-60.

Adamakos, H., Ryan, K., Ullman, D. G., Pascoe, J., Diaz, R., & Chessare, J. (1986). Maternal social support as a predictor of mother-child stress and stimulation. *Child Abuse and Neglect, 10*, 463-470.

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.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369-406.

Arbuckle, J. L. (2005). *Amos (Version 6.0)* [Computer Program]. Chicago: SPSS.

Arbuckle, J. L. (2008). *Amos (Version 17.0.0)* [Computer program]. Chicago: SPSS.

Azuma, H., & Kashiwagi, K. (1987). Descriptors for an intelligent person: A Japanese study. *Japanese Psychological Research, 29*, 17-26.

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.

Baker-Henningham, H., Powell, C., Walker, S., & Grantham-McGregor, S. (2003). Mothers of undernourished Jamaican children have poorer psychosocial functioning and this is associated with stimulation provided in the home. *European Journal of Clinical Nutrition, 57*, 786-792.

Berry, J. W., Poortinga, Y. H., Segall, M. H., & Dasen, P. R. (2002). *Cross-cultural psychology: Research and applications*. Cambridge, United Kingdom: Cambridge University Press.

Bickley, P. G., Keith, T. Z., & Wolfle, L. M. (1995). The three-stratum theory of cognitive abilities: Test of the structure of intelligence across the life span. *Intelligence, 20*, 309-328.

Boivin, M. J., Chounramany, C., Giordani, B., Xaisida, S., Choulamountry, L., Pholsena, P., et al. (1996). Validating a cognitive ability testing protocol with Lao children for community development applications. *Neuropsychology, 4*, 588-599.

Born, M. P., Bleichrodt, N., & Van der Flier, H. (1987). Cross-cultural comparison of sex-related differences on intelligence tests. *Journal of Cross-Cultural Psychology, 18*, 283-314.

Bradley, R. H., Caldwell, B. M., Rock, S. L., Ramey, C. T., Barnard, K. E., Gray, C., et al. (1989). Home environment and cognitive development in the first 3 years of life: A collaborative study involving six sites and three ethnic groups in North America. *Developmental Psychology, 25*, 217-235.

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology, 53*, 371-399.

Bradley, R. H., & Corwyn, R. F. (2005). Caring for children around the world: A view from HOME. *International Journal of Behavioral Development, 29*, 468-478.

Bradley, R. H., Corwyn, R. F., & Whiteside-Mansell, L. (1996). Life at home: Same time, different places - An examination of the HOME Inventory in different cultures. *Early Development and Parenting, 4*, 251-269.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.

Brooks-Gunn, J., & Duncan, G. J. (1997). The effects of poverty on children. *The Future of Children, 7*, 55-71.

Caldwell, B. M., & Bradley, R. H. (1984). *Home Observation for Measurement of the Environment*. Little Rock, AR: University of Arkansas.

Caldwell, B. M., & Bradley, R. H. (2003). *HOME Inventory administration manual*. Little Rock, AR: Print Design, Inc.

Carraher, T. N., Carraher, D. W., & Schliemann, A. D. (1985). Mathematics in the streets and in schools. *British Journal of Developmental Psychology, 3*, 21-29.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, United Kingdom: Cambridge University Press.

Chao, R., & Tseng, V. (2002). Parenting of Asians. In M. H. Bornstein (Ed.), *Handbook of parenting (second edition), volume 4: Social conditions and applied parenting* (pp. 59-93). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Cicchetti, D., Bronen, R., Spencer, S., Haut, S., Berg, A., Oliver, P., et al. (2006). Rating scales, scales of measurement, issues of reliability: Resolving some critical issues for clinicians and researchers. *The Journal of Nervous and Mental Disease, 194*, 557-564.

Cohen, S., Gottlieb, B. H., & Underwood, L. G. (2000). Social relationships and health. In S. Cohen, L. G. Underwood, & B. H. Gottlieb (Eds.), *Social support measurement and intervention* (pp. 3-25). Oxford, United Kingdom: Oxford University Press.

Conant, L. L., Fastenau, P. S., Giordani, B., Boivin, M. J., Opel, B., & Nseyila, D. D. (1999). Modality specificity of memory span tasks among Zairian children: A developmental perspective. *Journal of Clinical and Experimental Neuropsychology, 21*, 375-384.

Conger, R. D., Wallace, L. E., Sun, Y., Simons, R. L., McLoyd, V. C., & Brody, G. H. (2002). Economic pressure in African American families: A replication and extension of the family stress model. *Developmental Psychology, 38*, 179-193.

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*, 163-183.

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.

Davis-Kean, P. E. (2005). The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. *Journal of Family Psychology, 19*, 294-304.

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.

Demetriou, A., Shayer, M., & Efklides, A. (1992). *Neo Piagetian theories of cognitive development*. London: Routledge.

Dolan, C. V., Roorda, W., & Wicherts, J. M. (2004). Two failures of Spearman's hypothesis: The GATB in Holland and the JAT in South Africa. *Intelligence, 32*, 155-173.

Duncan, G. J., Brooks-Gunn, J., & Klebanov, P. K. (1994). Economic deprivation and early childhood development. *Child Development, 65*, 296-318.

Dunn, O. J., & Clark, V. A. (1969). Correlation coefficients measured on the same individuals. *Journal of the American Statistical Association, 64*, 366-377.

Durand, M., Hulme, C., Larkin, R., & Snowling, M. (2005). The cognitive foundations of reading and arithmetic skills in 7- to 10-year-olds. *Journal of Experimental Child Psychology, 91*, 113-136.

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.

Evans, G. W., Boxhill, L., & Pinkava, M. (2008). Poverty and maternal responsiveness: The role of maternal stress and social resources. *International Journal of Behavioral Development, 32*, 232-237.

Fairweather, H. (1976). Sex differences in cognition. *Cognition, 4*, 231-280.

Filmer, D., & Pritchett, L. H. (2001). Estimating wealth effects without expenditure data - or tears: An application to educational enrollments in states of India. *Demography, 38*, 115-132.

Flanagan, D. P., & Harrison, P. L. (Eds.). (2005). *Contemporary intellectual assessment: Theories, tests, and issues. Second edition*. New York: The Guilford Press.

The free dictionary. (n.d.). Retrieved May 14th, 2009, from http://www.thefreedictionary.com/cognition

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*, 177-190.

Gaunekar, G., Patel, V., & Rane, A. (2005). The impact and patterns of hazardous drinking amongst male industrial workers in Goa, India. *Social Psychiatry and Psychiatric Epidemiology, 40*, 267-275.

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., Van de Vijver, F. J. R., Weiss, L. G., & Saklofske, D. H. (2003). A cross-cultural analysis of the WISC-III. In J. Georgas, L. G. Weiss, F. J. R. Van de Vijver, & D. H. Saklofske (Eds.), *Culture and children's intelligence: Cross-cultural analysis of the WISC-III* (pp. 277-313). San Diego, CA: Academic Press.

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.

Gershoff, E. T. (2002). Corporal punishment by parents and associated child behaviors and experiences: A meta-analytic and theoretical review. *Psychological Bulletin, 128*, 539-579.

Gershoff, E. T., Aber, J. L., Raver, C. C., & Lennon, M. C. (2007). Income is not enough: Incorporating material hardship into models of income associations with parenting and child development. *Child Development, 78*, 70-95.

Goldberg, D. P. (1978). *Manual of the General Health Questionnaire*. Windsor: NFER.

Goldberg, D. P., Gater, R., Sartorius, N., Ustun, T. B., Piccinelli, M., Gureje, O., et al. (1997). The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychological Medicine, 27*, 191-197.

Goodchild, M. E., & Duncan-Jones, P. (1985). Chronicity and the General Health Questionnaire. *The British Journal of Psychiatry, 146*, 55-61.

Grigorenko, E. L., Geissler, P. W., Prince, R., Okatcha, F., Nokes, C., Kenny, D. A., et al. (2001). The organisation of Luo conceptions of intelligence: A study of implicit theories in a Kenyan village. *International Journal of Behavioral Development, 25*, 367-378.

Guo, G., & Harris, K. M. (2000). The mechanisms mediating the effects of poverty on children's intellectual development. *Demography, 37*, 431-447.

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.

Harkness, S., & Super, C. M. (2002). Culture and parenting. In M. H. Bornstein (Ed.), *Handbook of parenting (second edition), volume 2: Biology and ecology of parenting* (pp. 253-280). Mahwah, NJ: Lawrence Erlbaum Associates.

Hartmann, P., Kruuse, N. H. S., & Nyborg, H. (2007). Testing the cross-racial generality of Spearman's hypothesis in two samples. *Intelligence, 35*, 47-57.

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.

Hessels, M. G. P. (1997). Low IQ but high learning potential: Why Zeyneb and Moussa do not belong in special education. *Educational and Child Psychology, 14*, 121-136.

Hoff, E., Laursen, B., & Tardif, T. (2002). Socieconomic status and parenting. In M. H. Bornstein (Ed.), *Handbook of parenting (second edition), volume 2: Biology and ecology of parenting* (pp. 231-252). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

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 Disctrict, Kenya. *Journal of the International Neuropsychological Society, 10*, 246-260.

Hunter, W. M., Jain, D., Sadowski, L. S., & Sanhueza, A. I. (2000). Risk factors for severe child discipline practices in rural India. *Journal of Pediatric Psychology, 25*, 435-447.

Hutton, U. M. Z., & Towse, J. N. (2001). Short-term memory and working memory as indices of children's cognitive skills. *Memory, 9*, 383-394.

International Institute for Population Sciences (IIPS) and ORC Macro. (2000). *National Family Health Survey (NFHS-2), 1998-99: India*. Mumbai: IIPS.

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.

Jansen, P., & Greenop, K. (2008). Factor analyses of the Kaufman Assessment Battery for Children assessed longitudinally at 5 and 10 years. *South African Journal of Psychology, 38*, 355-365.

Jensen, A. R. (1985). The nature of the Black-White difference on various psychometric tests: Spearman's hypothesis. *Behavioral and Brain Sciences, 8*, 193-263.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.

Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman Assessment Battery for Children*. Circle Pines, MN: AGS Publishing.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children, Second Edition: Manual*. Circle Pines, MN: AGS Publishing.

Kaufman, J. C., Kaufman, A. S., Kaufman-Singer, J., & Kaufman, N. L. (2005). The Kaufman Assessment Battery for Children-Second Edition and the Kaufman Adolescent and Adult Intelligence Test. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 344-370). New York: The Guilford Press.

Keane, M., & Eysenck, M. W. (2005). *Cognitive psychology: A student's handbook*. Hove, United Kingdom: Psychology Press.

Kohs, S. C. (1927). *Intelligence measurement*. New York: Macmillan.

Krishnakumar, A., & Black, M. M. (2002). Longitudinal predictors of competence among African American children: The role of distal and proximal risk factors. *Applied Developmental Psychology, 23*, 237-266.

Kroeber, A. L., & Kluckhohn, C. (1963). *Culture: A critical review of concepts and definitions*. Cambridge, MA: Harvard University.

Lansford, J. E., Chang, L., Dodge, K. A., Malone, P. S., Oburu, P., Palmérus, K., et al. (2005). Physical discipline and children's adjustment: Cultural normativeness as a moderator. *Child Development, 76*, 1234-1246.

Lugo-Gil, J., & Tamis-LeMonda, C. S. (2008). Family resources and parenting quality: Links to children's cognitive development across the first 3 years. *Child Development, 79*, 1065-1085.

Lynn, R., & Owen, K. (1994). Spearman's hypothesis and test score differences between Whites, Indians, and Blacks in South Africa. *Journal of General Psychology, 121*, 27-36.

Macionis, J. J., & Plummer, K. (1998). *Sociology: A global introduction*. London: Prentice Hall Europe.

Magnuson, K. A., & Duncan, G. J. (2002). Parents in poverty. In M. H. Bornstein (Ed.), *Handbook of parenting (second edition), volume 4: Social conditions and applied parenting* (pp. 95-121). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Malda, M., Van de Vijver, F. J. R., Srinivasan, K., Transler, C., & Sukumar, P. (2009). *Stimulation in the home positively relates to children's cognitive performance in Bangalore, India.* Manuscript submitted for publication.

Malda, M., Van de Vijver, F. J. R., Srinivasan, K., Transler, C., & Sukumar, P. (in press). Traveling with cognitive tests: Testing the validity of a KABC-II adaptation in India. *Assessment*.

Malda, M., Van de Vijver, F. J. R., Srinivasan, K., Transler, C., Sukumar, P., & Rao, K. (2008). Adapting a cognitive test for a different culture: An illustration of qualitative procedures. *Psychology Science Quarterly, 50*, 451-468.

Malda, M., Van de Vijver, F. J. R., & Temane, M. Q. (2009). *Rugby versus soccer in South Africa: Content familiarity explains most cross-cultural differences in cognitive test scores.* Manuscript submitted for publication.

Mardell-Czudnowski, C. (1995). Performance of Asian and White children on the K-ABC: Understanding information processing differences. *Assessment, 2*, 19-29.

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. (1990). The impact of economic hardship on black families and children: Psychological distress, parenting, and socioemotional development. *Child Development, 61*, 311-346.

McLoyd, V. C. (1998). Socioeconomic disadvantage and child development. *American Psychologist, 53*, 185-204.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: American Council on Education and Macmillan Publishing Company.

Mishra, R. C. (1997). Cognition and cognitive development. In J. W. Berry, P. R. Dasen, & T. S. Saraswathi (Eds.), *Handbook of cross-cultural psychology (Vol. 2): Basic processes and human development* (pp. 143-175). Needham Heights, MA: Allyn & Bacon.

Misra, G., Sahoo, F. M., & Puhan, B. N. (1997). Cultural bias in testing: India. *European Review of Applied Psychology, 47*, 309-316.

Mohan, D., Chopra, A., Ray, R., & Sethi, H. (2000). Alcohol consumption in India: A cross-sectional study. In *Surveys of drinking patterns and problems in seven developing countries* (pp. 103-113). World Health Organization: Department of Mental Health and Substance Dependence. Retrieved January 26, 2009, from http://www.unicri.it/min.san.bollettino/dati/AlcBrochur.pdf

Mohan, D., Desai, N. G., Chopra, A., & Sethi, H. (1992). Rapid survey of substance abuse disorders on urban slums of New Delhi. *Indian Journal of Medical Research, 96*, 122-127.

Mohan, I. (1998). India. In M. Grant (Ed.), *Alcohol and emerging markets: Patterns, problems and responses* (pp. 145-175). Philadelphia, PA: Psychology Press.

Moon, S., McLean, J. E., & Kaufman, A. S. (2003). A cross-cultural validation of the sequential-simultaneous theory of intelligence in children. *School Psychology International, 24*, 449-461.

Moos, R. H., Insel, P. M., & Humphrey, B. (1974). *Manual for the Family Environment Scale*. Palo Alto, CA: Consulting Psychologists Press.

Nair, M. K. C., & Radhakrishnan, R. S. (2004). Early childhood development in deprived urban settlements. *Indian Pediatrics, 41*, 227-237.

Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity - facets of a cognitive ability construct. *Personality and Individual Differences, 29*, 1017-1045.

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.

Pandey, J. (2006). India. In J. Georgas, J. W. Berry, F. J. R. Van de Vijver, C. Kağitçibaşi, & Y. H. Poortinga (Eds.), *Families across cultures: A 30-nation psychological study* (pp. 362-369). Cambridge, United Kingdom: Cambridge University Press.

Paolucci, E. O., & Violato, C. (2004). A meta-analysis of the published research on affective, cognitive, and behavioral effects of corporal punishment. *The Journal of Psychology, 138*, 197-221.

Pascoe, J. M., & Earp, J. A. (1984). The effects of mothers' social support and life changes on the stimulation of their children in the home. *American Journal of Public Health, 74*, 358-360.

Petterson, S. M., & Albers, A. B. (2001). Effects of poverty and maternal depression on early child development. *Child Development, 72*, 1794-1813.

Pothen, M., Kuruvilla, A., Philip, K., Joseph, A., & Jacob, K. S. (2003). Common mental disorders among primary care attenders in Vellore, South India: Nature, prevalence and risk factors. *International Journal of Social Psychiatry, 49*, 119-125.

Ramaa, S. (1994). *Arithmetic diagnostic test for primary school children*. Mysore, India: Chethana Book House.

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.

Reynolds, M. R., Keith, T. Z., Goldenring-Fine, J., Fisher, M. E., & Low, J. A. (2007). Confirmatory factor structure of the Kaufman Assessment Battery for Children-Second edition: Consistency with Cattell-Horn-Carroll theory. *School Psychology Quarterly, 22*, 511-539.

Reynolds, M. R., Keith, T. Z., Ridley, K. P., & Patel, P. G. (2008). Sex differences in latent general and broad cognitive abilities for children and youth: Evidence from higher-order MG-MACS and MIMIC models. *Intelligence, 36*, 236-260.

Saraswathi, T. S., & Dutta, R. (1988). Current trends in developmental psychology: A life span perspective. In J. Pandey (Ed.), *Psychology in India: The state-of-the-art. Volume 1: Personality and mental processes* (pp. 93-152). New Delhi, India: Sage Publications.

Schweizer, K., & Moosbrugger, H. (2004). Attention and working memory as predictors of intelligence. *Intelligence, 32*, 329-347.

Segall, M. H., Dasen, P. R., Berry, J. W., & Poortinga, Y. H. (1999). *Human behavior in global perspective: An introduction to cross-cultural psychology*. Boston: Allyn & Bacon.

Serpell, R. (1979). How specific are perceptual skills? A cross-cultural study of pattern reproduction. *British Journal of Psychology, 70*, 365-380.

Serpell, R. (1993). *The significance of schooling: Life-journeys in an African society*. Cambridge, United Kingdom: Cambridge University Press.

Sinha, D. (1988). Basic Indian values and behaviour dispositions in the context of national development: An appraisal. In D. Sinha & H. S. R. Kao (Eds.), *Social values and development: Asian perspectives* (pp. 31-55). New Delhi, India: Sage Publications.

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.

Snijders, J. T., Tellegen, P. J., & Laros, J. A. (1989). *Snijders-Oomen Non-verbal intelligence test: SON-R 5,5-17. Manual and research report*. Groningen, the Netherlands: Wolters-Noordhoff.

Sriram, T. G., Chandrashekar, C. R., Isaac, M. K., & Shanmugham, V. (1989). The General Health Questionnaire (GHQ): Comparison of the English version and a translated Indian version. *Social Psychiatry and Psychiatric Epidemiology, 24*, 317-320.

Srivastava, A. K., & Misra, G. (2001). Lay people's understanding and use of intelligence: An Indian perspective. *Psychology and Developing Societies, 13*, 25-49.

Stansfield, C. W. (2003). Test translation and adaptation in public education in the USA. *Language Testing, 20*, 189-207.

Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic testing: The nature and measurement of learning potential*. Cambridge, United Kingdom: Cambridge University Press.

Sternberg, R. J., Grigorenko, E. L., Ngorosho, D., Tantufuye, E., Mbise, A., Nokes, C., et al. (2002). Assessing intellectual potential in rural Tanzanian school children. *Intelligence, 30*, 141-162.

Süß, H.-M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working - memory capacity explains reasoning ability - and a little bit more. *Intelligence, 30*, 261-288.

Swanson, H. L. (2008). Working memory and intelligence in children: What develops? *Journal of Educational Psychology, 100*, 581-602.

Tardy, C. H. (1985). Social support measurement. *American Journal of Community Psychology, 13*, 187-202.

Taub, G. E., & McGrew, K. S. (2004). A confirmatory factor analysis of Cattell-Horn-Carroll theory and cross-age invariance of the Woodcock-Johnson Tests of Cognitive Abilities III. *School Psychology Quarterly, 19*, 72-87.

Te Nijenhuis, J., & Van der Flier, H. (1997). Comparability of GATB scores for immigrants and majority group members: Some Dutch findings. *Journal of Applied Psychology, 82*, 675-687.

Thurstone, L. L. (1931). *The reliability and validity of tests: Derivation and interpretation of fundamental formulae concerned with reliability and validity of tests and illustrative problems*. Ann Harbor, MI: Edwards Brothers.

Tylor, E. B. (1871). *Primitive culture (2 Vols.)*. London: Murray.

Tzuriel, D. (2001). *Dynamic assessment of young children*. New York: Kluwer Academic / Plenum Publishers.

Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between Operation Span and Raven. *Intelligence, 33*, 67-81.

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.

Verney, S. P., Granholm, E., Marshall, S. P., Malcarne, V. L., & Saccuzzo, D. P. (2005). Culture-fair cognitive ability assessment: Information processing and psychophysiological approaches. *Assessment, 12*, 303-319.

Vock, M., & Holling, H. (2008). The measurement of visuo-spatial and verbal-numerical working memory: Development of IRT-based scales. *Intelligence, 36*, 161-182.

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.

Wasserman, J. D., & Tulsky, D. S. (2005). A history of intelligence assessment. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 3-22). New York: The Guilford Press.

Wechsler, D. (1939). *The measurement of adult intelligence*. Baltimore: Williams & Wilkins.

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: 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.

Whiting, B. (1976). The problem of the packaged variable. In K. Riegel & J. Meacham (Eds.), *The developing individual in a changing world* (Vol. 1, pp. 303-309). The Hague: Mouton.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement equivalence of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention* (pp. 281-324). Washington, DC: American Psychological Association.

Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage Publications.

Wills, T. A., & Shinar, O. (2000). Measuring perceived and received social support. In S. Cohen, L. G. Underwood, & B. H. Gottlieb (Eds.), *Social support measurement and intervention* (pp. 86-135). Oxford, United Kingdom: Oxford University Press.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III*. Itasca, IL: Riverside.

Yeung, W. J., Linver, M. R., & Brooks-Gunn, J. (2002). How money matters for young children's development: Parental investment and family processes. *Child Development, 73*, 1861-1879.

Samenvatting (Summary in Dutch)

Zoals het klokje thuis tikt, tikt het nergens: Over de relatie tussen cultuur en cognitie van kinderen

In dit proefschrift bestudeer ik de relatie tussen cultuur en cognitie van kinderen. Kinderen uit verschillende culturen scoren vaak verschillend op cognitieve tests. De oorzaak van deze bevinding zou kunnen liggen in verschillen in cognitieve vaardigheden, maar ook in de mate waarin de tests cultuurgebonden zijn. Kinderen scoren hoger op tests naarmate deze meer op hun eigen cultuur zijn afgestemd. Als de cultuurgebondenheid van cognitieve tests groot is, vertonen kinderen uit andere culturen lagere scores vanwege hun relatieve onbekendheid met de (inhoud van de) tests en niet per se vanwege een gebrek aan cognitieve vaardigheden. Er is dan sprake van bias; bias is een verzamelterm voor alle factoren die de vergelijkbaarheid van metingen verkregen bij verschillende (culturele) groepen bedreigen. Bias-gerelateerde problemen bij het meten van cognitie en bij het meten van kenmerken van de thuisomgeving van een kind spelen een belangrijke rol bij het bestuderen van de relatie tussen de concepten cultuur en cognitie. In dit proefschrift onderzoek ik de link tussen deze concepten vanuit drie invalshoeken waarbij ik rekening houd met bias. De eerste invalshoek heeft betrekking op het aantonen en reduceren van bias in cognitieve tests (Hoofdstuk 2 en 3); de tweede op de relatie tussen de thuisomgeving en cognitie van een kind (Hoofdstuk 4), de derde op het manipuleren van bias om antwoord te geven op de vraag waarom sommige tests grotere scoreverschillen tussen culturen laten zien dan andere (Hoofdstuk 5).

Dit proefschrift levert op drie manieren een bijdrage aan de huidige literatuur. Ten eerste integreert het de concepten cultuur, cognitie en bias. Bias wordt niet slechts gezien als een meetprobleem dat vermeden moet worden; het dient standaard in het design van een studie te worden opgenomen. Dit doe ik in dit proefschrift door op bias te anticiperen, het te bestuderen en het te manipuleren.

Door rekening te houden met bias in plaats van te proberen het te vermijden is een beter zicht te krijgen op de relatie tussen cultuur en cognitie. Ten tweede bekijkt dit proefschrift deze relatie niet vanuit één invalshoek, maar vanuit meerdere invalshoeken (zoals eerder beschreven). Ten derde wordt de link tussen cultuur en cognitie bestudeerd in twee daarvoor bij uitstek geschikte contexten, namelijk India en Zuid-Afrika. Beide landen zijn multicultureel en behoren niet tot de westerse landen waar de meeste cognitieve tests worden ontwikkeld. Veel tests die gebruikt worden in India en Zuid-Afrika komen uit deze westerse landen. Hoe groter de crossculturele verschillen zijn tussen de oorspronkelijke context van een test en een andere context, des te meer potentiële bronnen van bias er zijn. Grote cultuurverschillen leveren dan ook goede condities voor het uitvoeren van een kritische test van de (crossculturele) toepasbaarheid van cognitieve tests en van modellen die de leefomgeving van een kind relateren aan diens cognitieve prestaties.

Testadaptatie houdt in dat een test die in een bepaalde taal en vanuit een bepaalde culturele achtergrond is ontwikkeld aan een andere taal en culturele achtergrond wordt aangepast. Hoofdstuk 2 beschreef een kwalitatieve procedure om cognitieve tests te adapteren, bestaande uit iteraties van vertalen, piloten, en het aanpassen van het instrument. Ik maakte een onderscheid tussen vijf soorten testadaptaties; aanpassingen kunnen gebaseerd zijn op, respectievelijk, het onderliggende construct, taal aspecten, culturele gebruiken, theorieën die ten grondslag liggen aan de tests, of de bekendheid van stimuli en test procedures. In Hoofdstuk 2 werden met name adaptaties op basis van onderliggende theorieën en bekendheid met stimuli uitgevoerd. De beschreven procedure werd toegepast om de Kaufman Assessment Battery for Children, second edition (KABC-II) geschikt te maken voor 6 tot 10 jaar oude Kannada sprekende kinderen van lage sociaal-economische status in Bangalore (India). Uitgebreide adaptaties waren nodig voor iedere subtest; de adaptaties betroffen de testinstructies, de iteminhoud van zowel verbale als niet-verbale tests en de itemvolgorde. Er werd geconcludeerd dat deze kwalitatieve benadering adequaat bleek voor het identificeren van allerlei problemen met de toepasbaarheid van de KABC-II voor deze steekproef die onopgemerkt zouden zijn gebleven bij een directe vertaling van het originele instrument. De studie toonde aan dat voor een adequaat gebruik van westerse cognitieve instrumenten in een niet-westerse context een nauwkeurige analyse van hun toepasbaarheid nodig is.

Hoofdstuk 3 evalueerde de psychometrische kwaliteit van de adaptatie van de KABC-II die in het vorige hoofdstuk werd beschreven. Data van 598 kinderen lieten hoge betrouwbaarheden zien voor alle subtests, het Cattell-Horn-Carroll model dat ten grondslag ligt aan de originele KABC-II werd grotendeels gerepliceerd, en correlaties met demografische eigenschappen en schoolprestaties waren in lijn met de verwachtingen. De subtests laadden relatief hoog op een algemene cognitieve factor, waarschijnlijk vanwege de onbekendheid en daardoor de hoge complexiteit van deze tests voor deze kinderen. De bevindingen bevestigden de geschiktheid en validiteit van de KABC-II adaptatie. De resultaten van Hoofdstuk 2 en 3 geven tezamen aan dat testadaptaties slechts adequaat kunnen zijn als ze zowel aan kwalitatieve als kwantitatieve (statistische) criteria voldoen.

Hoofdstuk 4 bestudeerde de relatie tussen de thuisomgeving en cognitieve prestaties van Indiase kinderen. De primaire verzorgers van 532 kinderen waarbij de KABC-II was afgenomen (zie Hoofdstuk 3) werden geïnterviewd. Een padmodel bevestigde de toepasbaarheid van omgeving-cognitie relaties die in de literatuur voorgesteld zijn; ik doel hier op het investeringsmodel en het familie stressmodel, die beide voornamelijk op westerse data gebaseerd zijn. Meer proximale variabelen (de geestelijke gezondheid van de primaire verzorger, sociale steun, gezinsconflicten, en gedragingen van de ouders) lieten sterkere associaties zien met cognitieve prestaties dan meer distale variabelen (sociaal-economische status); de link tussen laatstgenoemde en cognitieve uitkomsten werd volledig gemedieerd door de meer proximale variabelen. Alleen de betrokkenheid van de ouders bij hun kind door middel van het bieden van directe stimulatie en structuur was direct gerelateerd aan uitkomsten. Interventies die erop gericht zijn om de ontwikkeling van kinderen te bevorderen zouden zich daarom moeten richten op de meest proximale gedragingen van de ouders.

Ik veronderstel in Hoofdstuk 5 dat crossculturele verschillen in cognitieve test scores afhankelijk zijn van de culturele complexiteit van een test (Culturele Complexiteits-Hypothese) en niet zozeer van de cognitieve complexiteit ervan (Spearman's Hypothese). Culturele complexiteit werd in dit hoofdstuk geconceptualiseerd als de mate waarin de testinhoud (woorden en plaatjes) bekender is voor één van de te vergelijken groepen. In dit hoofdstuk werd deze bekendheid van testinhoud gemanipuleerd voor vijf cognitieve tests (voor korte termijngeheugen, aandacht, werkgeheugen, en figuraal en verbaal redeneren) om het effect ervan op testprestatie te bestuderen. Van iedere test werden twee versies gemaakt: een Afrikaanse en een Tswana testversie, met een inhoud die erg bekend is voor Zuid-Afrikaanse kinderen met respectievelijk een Afrikaanse en Tswana achtergrond. De steekproef bestond uit 501 schoolgaande kinderen uit de derde en vierde klas van het basisonderwijs (161 urbane Afrikaans sprekende kinderen, en 181 urbane en 159 rurale Setswana sprekende kinderen). De resultaten waren niet in overeenstemming met Spearman's Hypothese en leverden bevestiging voor de Culturele Complexiteits-Hypothese; kinderen presteerden over het algemeen beter op de testversie die voor hun eigen groep was ontwikkeld dan op de andere testversie. Dit effect was het sterkst op de cognitief (en cultureel) complexe werkgeheugen en redeneertests. De bekendheid van de inhoud van een test is een belangrijke moderator van cognitieve testprestaties waar, met name bij crossculturele vergelijkingen van testscores, rekening mee gehouden dient te worden.

De praktische relevantie van dit proefschrift komt tot uitdrukking in drie aspecten. Ten eerste vergroot het ons inzicht in de aard van crossculturele scoreverschillen. Ik heb in Hoofdstuk 5 laten zien dat culturele complexiteit een belangrijkere rol speelt in het verklaren van scoreverschillen dan het niveau van cognitieve vaardigheden dat ervoor nodig was om op een test goed te presteren. Onderzoek waarvan het doel is om crossculturele verschillen in testscores onder de loep te nemen, moet rekening houden met beide verklaringen. We moeten voorzichtig zijn met het trekken van conclusies over de vraag of cognitieve of culturele complexiteit belangrijker is.

Ten tweede levert het proefschrift richtlijnen voor het ontwikkelen van een test die adequaat is voor een bepaalde cultuur. Het is gebruikelijk om tests direct te vertalen in de taal van de doelgroep zonder verdere aanpassingen te maken. Wanneer tests in een andere context worden gebruikt zonder veranderingen aan te brengen in stimuli, instructies, of procedures, is de validiteit twijfelachtig in het geval van een grote culturele kloof tussen de originele cultuur en de cultuur van de doelgroep. Dit proefschrift laat zien dat adaptatie een bruikbaar middel kan zijn om deze kloof te overbruggen. Het is echter onmogelijk om een test te maken die even bekend is voor alle bestudeerde groepen, wat impliceert dat het onmogelijk is om bias volledig te verwijderen. Blijkbaar levert decontextualiseren niet het gewenste effect en is het beter om juist tests te contextualiseren (i.e., rekening houden met bias in plaats van proberen het te vermijden), door ervoor te zorgen dat tests zo bekend als mogelijk zijn voor de doelgroep.

Een geschikt instrumentarium om het functioneren van een kind te meten is nodig om betekenisvolle interventies uit te voeren ten behoeve van de ontwikkeling van een kind. Om doelen te identificeren voor dergelijke interventies is het van belang te weten welke factoren in het dagelijks leven van een kind positief of negatief samenhangen met deze ontwikkeling. Dit proefschrift draagt bij aan dit proces door de (crossculturele validiteit van) relaties te bestuderen tussen bepaalde variabelen in the thuisomgeving en de cognitieve prestaties van een kind. Dit is de derde praktische bijdrage van dit proefschrift. In Hoofdstuk 4 heb ik laten zien dat de aspecten die ertoe doen, die aspecten in de thuisomgeving zijn die het meest proximaal zijn ten opzichte van deze prestaties. Moeders bewust maken van de directe invloed die zij kunnen hebben op de cognitieve prestaties van hun kinderen vormt een belangrijk doel van interventies ter stimulering van de ontwikkeling van kinderen. Directe betrokkenheid van de ouders bij het kind door het bieden van structuur en stimulatie blijkt essentieel.

Hoe zijn cultuur en de cognitie van kinderen aan elkaar gerelateerd? Er is geen eenvoudig antwoord te geven op de vraag die ten grondslag ligt aan dit proefschrift. Er is een relatie, dat is zeker. Cultuur bepaalt onder meer welke factoren in de thuisomgeving van een kind van belang zijn voor een goede ontwikkeling en welke cognitieve vaardigheden gewaardeerd (en daardoor gestimuleerd) worden. Cultuur beïnvloedt ook de mate waarin tests bias vertonen in het meten van deze omgevingsfactoren en cognitieve vaardigheden. Waar cultuur is, is bias. Het is niet mogelijk om helemaal van bias af te komen; vertrouwdheid met (de inhoud van) tests heeft een grote invloed op testprestaties. Het hangt echter af van de situatie in welke mate de aanwezigheid van bias een probleem is. Toch concludeer ik dat een (cognitieve) test het meest thuis is in het land of de context waarin deze is ontwikkeld of waarvoor deze is geadapteerd. Het is immers nergens zoals thuis.

Acknowledgements

Juist door de extremen die mijn promotieproject heeft gekend, heb ik de afgelopen jaren als zeer speciaal ervaren. Des te meer omdat ik in deze periode veel tijd met bijzondere mensen heb doorgebracht. Die mensen bedank ik hier dan ook graag.

Fons, dankzij jou ben ik in aanraking gekomen met crossculturele psychologie. Onze samenwerking was soepel en prettig en ik bewonder je snelheid en je flexibiliteit. Iedere keer als ik dacht even van een paper af te zijn, vond ik een "versie retour" in mijn inbox... Tijdens mijn bezoeken aan India en Zuid-Afrika was 's avonds of in het weekend skypen voor jou (vreemd genoeg) nooit een probleem. Bedankt ook dat je erop vertrouwde dat ik tijdens het veldwerk de juiste beslissingen zou nemen. Fons, cheers!

Mijn dank aan Unilever Food and Health Research Institute is groot. Dankzij hen ben ik in staat gesteld om van een tijdelijk project een promotieproject te maken. Jan-Willem, bedankt dat je mij deze kans hebt gegeven. Catherine, merci for your efforts, for your patience, for keeping me sharp (also when I did not feel like being sharp...), and for the personal touch you gave to our communication. Ans, wat ben ik blij dat wij onze India-ervaringen met elkaar konden delen.

Anura, thank you for welcoming me at St. John's Research Institute (Bangalore, India) so many times. Dr. Srinivasan, our collaboration was good and I thank you for adding your valuable perspective and knowledge to the CHAMPION study. Sumithra, when I think of my time in India, I think of you. Thanks for the precious conversations. Prathima, it was great to be part of this project together and it is even better to be your friend. Aruna, thanks for sharing your views on the many sides of life in India.

South Africa, my second home. Ian, thanks for helping me in creating an opportunity to return to South Africa in 2008. Michael, I immensely enjoyed our collaboration. Thank you for the great conversations and for your valuable input. Vera, thanks for your advice and care. Ansie, I greatly admire your drive and practical perspective and am very thankful for your help and kindness. Ina, you are the best! Dit is flippen hond lekker om saam met jou te kuier. Bennie and Carina, your kindness has made me feel at home in Potch. Charlotte, ons het en ons sal baie lekker gesels. Jaco, thanks for being you.

Judith, wat heb je geweldige tekeningen gemaakt!

I thank my colleagues from Cross-Cultural Psychology for accompanying me on the journey with destination PhD. Ype, bedankt voor je adviezen en met name voor de gesprekken in Zuid-Afrika. Sommige woorden heb ik echter wel met een korreltje zout genomen... Amina, thanks for bothering me. Special thanks to Elif; I had a really good and fun time with you, both in and out of the office.

En dan is er Sociale Psychologie. Jullie waren mijn maatjes en ik heb genoten van jullie gezelschap en de borrels. Marcel, bedankt. Marijke, ik heb altijd erg met (maar ook om) je gelachen. Lieven, je hebt mijn zwak voor het prachtige Vlaams vertienvoudigd. Marret, jouw "houd het hoofd koel en de voeten warm" neem ik met me mee. Yvette, ik denk dat ik nooit meer iemand ontmoet die bij haar volle verstand een kei-kolossale pauw op een kei-kleine kerstboom zet. Seger ("C"-g), je bent een garantie voor veel positiefs. Dankie voor je woorden, je betrokkenheid en je bewonderenswaardige interesse in alles om je heen.

Mijn vriendinnen Melissa, Esther vd W., Reineke, Ilona en Jouke, bedankt dat jullie er waren en er nog steeds zijn. Jullie haalden me regelmatig even uit de wetenschappelijke wolk en zetten me met beide benen op de grond. Esther H., fijn dat we met onze aio-ervaringen zo onvoorstelbaar parallel liepen; ik heb veel aan onze gesprekken gehad. Yaniv, thanks for caring and sharing, from wherever you are. Sas, bedankt voor het urenlange geklets en voor je inspirerende balans en flow. Inge (Inga), geweldige vriendinna, ik geniet van je enthousiasme en je toewijding. Marjos (Jossie), je bent er al heel lang voor me. Ik dank je voor je onmisbare vriendschap.

Kalle, mijn broer, bedankt voor je smsjes en ook voor de onverwachte bijdragen die je aan mijn proefschrift hebt geleverd. Als niet-psycholoog heb je met je nuchtere kijk ideeën aangedragen voor mijn India studie en heb je items voor de Zuid-Afrika studie bewerkt. Op deze manier reisde je altijd een beetje met me mee, en je bleek een fijne reisgenoot.

Pap en mam, ik hoor het jullie nog zeggen: "Wanneer ga je weer op reis? Dan horen we tenminste nog eens wat van je!". Ik belde jullie namelijk vaker op vanuit het buitenland dan vanuit Tilburg. Bedankt dat jullie me altijd mijn eigen ding lieten doen. Jullie hebben mijn spraakwatervallen (tot in den treure) aangehoord en hebben jullie "Miekje" aangemoedigd op de juiste momenten. Bedankt.

Cultuurpsychologen (b)lijken nogal eens ontheemd te zijn, geen thuis te hebben. Door negen keer af te reizen naar India en drie keer naar Zuid-Afrika heb ik aan het woord "thuis" een andere betekenis gegeven. Thuis wordt niet bepaald door een land of plaats; het zijn de mensen die het ergens tot een thuis maken. Mensen, bedankt!

Maike Malda
Tilburg, 1 augustus 2009