Tilburg University

Rugby versus soccer in South Africa

Malda, M.; van de Vijver, F.J.R.; Temane, Q.M.

Published in:
Intelligence

Publication date:
2010

Document Version
Publisher's PDF, also known as Version of record

Intelligence xxx (2010) xxx–xxx

Contents lists available at ScienceDirect
Intelligence

Rugby versus Soccer in South Africa: Content familiarity contributes to cross-cultural differences in cognitive test scores[☆]

Maike Malda [a,*], Fons J.R. van de Vijver [a,b], Q. Michael Temane [b]

[a] Tilburg University, the Netherlands
[b] North-West University, South Africa

ARTICLE INFO

Article history:
Received 16 July 2009
Received in revised form 24 July 2010
Accepted 28 July 2010
Available online xxxx

Keywords:
Cognition
Culture
Spearman's Hypothesis
Content familiarity
Validity
South Africa

ABSTRACT

In this study, cross-cultural differences in cognitive test scores are hypothesized to depend on a test's cultural complexity (Cultural Complexity Hypothesis: CCH), here conceptualized as 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, 181 urban, and 159 rural Tswana children (M_{age} = 9.37 years). Children generally performed best on the test version that was designed for their own group, particularly on the cognitively and culturally complex working memory and figural fluid reasoning tests. This relation between content familiarity and cognitive test performance supports CCH and disconfirms SH.

Cross-cultural differences in cognitive test scores are not well understood (Fagan & Holland, 2002, 2007, 2009; Helms-Lorenz, Van de Vijver, & Poortinga, 2003; Sternberg et al., 2002). 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, score differences between cultures can also be caused by test bias or by cultural differences in valued and therefore trained strategies to solve certain cognitive tasks (Bridgeman & Buttram, 1975; Montie & Fagan, 1988). In line with the option of test bias, Helms-Lorenz et al. (2003) state that a test's cognitive complexity could be confounded with its cultural complexity (which can cause test bias) and the latter may largely explain cross-cultural score differences. A test's cultural complexity refers to the extent to which specific cultural knowledge is required to perform well on this test, such as declarative (factual) knowledge and procedural knowledge (i.e., knowledge on cultural practices, such as sports or celebrations) that is shared in a particular culture. The extent to which test content (e.g., concepts, drawings) is more familiar to one of several compared cultural groups is a reflection of cultural complexity. 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

☆ We thank Ansie Kitching and Dipuo for their help during the pilot study. Our gratitude goes to Abegale, Andrea, Anita, Corlene, Hester, Ina-Mari, Kefilwe, Keitumetse, Kelebogile, Liesl, Lorato, Marguerite, Marichelle, Maryke, Masego, Myrtle, Segametsi, Sharlott, Surita, Tsholofelo, Victoria, and Vidette for administering the cognitive tests to the children. We would like to thank the principals and teachers of the participating schools for their interest and enthusiasm in our study. Last but not least, we thank all participating children.
* Corresponding author. Department of Psychology, Tilburg University, PO box 90153, 5000 LE Tilburg, the Netherlands. Tel.: +31 0 13 4662526.
E-mail address: maikemalda@gmail.com (M. Malda).

0160-2896/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2010.07.004

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 or Tswana school-age children in South Africa. Both versions were administered to children of both groups.

1. Cultural complexity

Successful performance on a test with a high cultural complexity requires specific cultural knowledge (Helms-Lorenz et al., 2003). 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 network facilitates the 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 is more familiar to one of the compared groups.

2. 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 Fig. 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 Fig. 1, the abilities become cognitively more complex.

3. Cognitive versus cultural complexity

Relatively 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 such tasks 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, ranging from batteries of reaction time tasks in which cognitive complexity was varied by increasing the number of response alternatives (Jensen, 1993) to broad cognitive batteries involving a range of cognitive abilities (reviews can be found in Jensen, 1985, 1998). Both Spearman and Jensen focused on IQ differences between Blacks and Whites in the United States; other researchers, testing SH among other groups in other contexts, using various cognitive tests, also found support for the hypothesis (e.g., Hartmann, Kruuse, & Nyborg, 2007; Lynn & Owen, 1994; Rushton, 2002; Te Nijenhuis, Evers, & Mur, 2000; Te Nijenhuis, Tolboom, Resing, & Bleichrodt, 2004; Te Nijenhuis & Van der Flier, 1997, 2003, 2004, 2005).

SH has met with both statistical and conceptual criticism. The adequacy of widely employed statistical procedures to test the hypothesis has 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). Also, Wicherts and Dolan (2010) found that the allegedly small cultural bias reported by Te Nijenhuis et al. (2004) appeared to be more substantial when re-analyzing their data by including a test for the equality of measurement intercepts over groups. In a similar vein, Wicherts and Johnson (2009) provided statistical arguments against the procedures used by Rushton (2002; see also Rushton, Bons, Vernon, & Cvorovic, 2007) to draw conclusions about the nature of group differences in cognitive abilities in terms of SH and the heritability of these differences.

Please cite this article as: Malda, M., et al., Rugby versus Soccer in South Africa: Content familiarity contributes to cross-cultural differences in cognitive test scores, *Intelligence* (2010), doi:10.1016/j.intell.2010.07.004

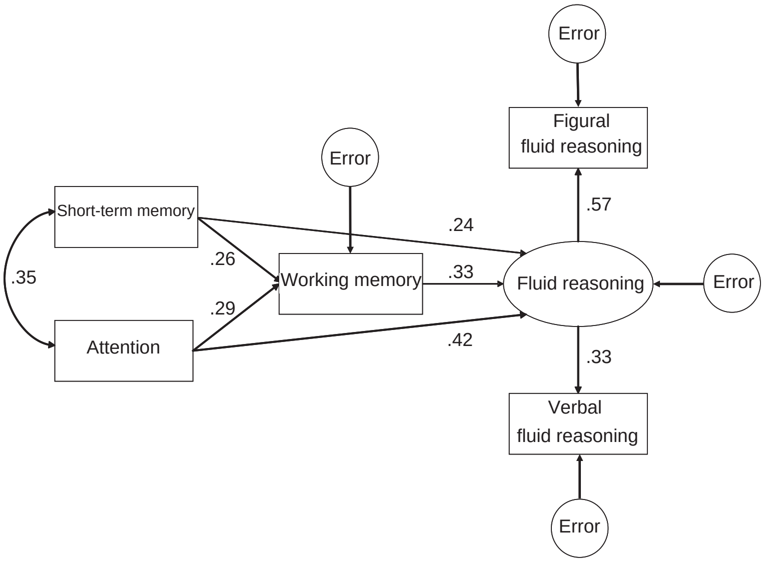

Fig. 1. Cognitive structure (hypothesized and confirmed). *Note*: all depicted parameter estimates are significant, $p<.01$.

There are substantive reasons to doubt the importance of cognitive complexity in explaining cross-cultural score differences. Hunt and Carlson (2007) as well as Wicherts, Dolan, Carlson, and Van der Maas (2010) mentioned environmental aspects as possible sources of group differences in cognitive test scores. The related concepts of affluence and socioeconomic status are suggested to contribute to these score differences (Van de Vijver, 1997). Also, the content of a test or the medium in which a test is administered has 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). Fagan and Holland (2002, 2007, 2009) have shown that majority and minority adults in the US obtained similar scores on cognitive tasks when they were equally exposed to the information required to successfully complete the tasks.

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). The model holds that cross-cultural score differences increase with the cultural complexity of the stimulus content (i.e., content familiarity). Cultural complexity can vary both across tests that measure different skills and across tests that measure a single skill using stimuli that differ in familiarity. To start with the former, tests that address simple information processing usually show less cultural complexity than tests addressing complex information processing (Helms-Lorenz et al., 2003; Vock & Holling, 2008; cf. however Jensen, 1993, and Te Nijenhuis & Van der Flier, 2003). 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 more complex processes, such as working memory, and will be largest on the most complex tests, such as fluid reasoning tests, which often require extensive cultural information to solve them. SH would predict the same increase in cross-cultural differences across these different types of tests, but due to an increase in *cognitive* rather than *cultural* complexity; hence, the common confounding of these two types of complexity. Therefore, the present study examines cross-cultural differences also across tests that measure the same skill using stimuli of different familiarity, as explained in the next section.

4. Present study and hypotheses

A test was constructed for each of the abilities displayed in Fig. 1. There were two versions of each test. One version contained items with a relatively higher content familiarity

Table 1
Expected score pattern for hypotheses.

Hypothesis	Test version	Expected score pattern Group
1 (CCH)	Afrikaans-culture	Afrikaans ≥ Rural Tswana
	Tswana-culture	Rural Tswana ≥ Afrikaans
2 (SH)	Afrikaans-culture	One group ≥ other groups
	Tswana-culture	The same group ≥ other groups
3	Afrikaans-culture	Urban Tswana in between other groups
	Tswana-culture	Urban Tswana in between other groups

Please cite this article as: Malda, M., et al., Rugby versus Soccer in South Africa: Content familiarity contributes to cross-cultural differences in cognitive test scores, *Intelligence* (2010), doi:10.1016/j.intell.2010.07.004

Table 2
Sample size for each group and test version.

Group	Test version									
	Afrikaans-culture					Tswana-culture				
	Total	Female	Male	Grade 3	Grade 4	Total	Female	Male	Grade 3	Grade 4
Afrikaans	80	39	41	40	40	81	41	40	41	40
Urban Tswana	91	38	53	44	47	90	46	44	44	46
Rural Tswana	82	39	43	42	40	77	42	35	41	36

for the Afrikaans than for the Tswana children, labeled the Afrikaans-culture test version. The other contained items with a relatively higher content familiarity for the Tswana than for the Afrikaans children, labeled the Tswana-culture test 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).

According to CCH, differences in test scores between the Afrikaans and Tswana children are due to differences in the content familiarity of the tests; CCH predicts that one group scores at least as high as the other group(s) on the test version developed for its own culture.[1] According to SH, differences between the Afrikaans and Tswana children are due to differences in the underlying cognitive abilities that are required to successfully complete the tests; SH predicts that one group scores consistently higher than the other(s) on both test versions (apart from random fluctuations). Generally, we expect the score differences to be largest between the Afrikaans and the rural Tswana group. Based on these two different lines of reasoning, two contradicting hypotheses can be formulated of which we expect to confirm the first (CCH) and disconfirm the second (SH; see also Table 1 for an overview of the expected score patterns):

1. CCH: Afrikaans children score at least as high as rural Tswana children on the Afrikaans-culture test version *and* rural Tswana children score at least as high as Afrikaans children on the Tswana-culture test version (differences between groups are smallest for short-term memory and attention, larger for working memory, and largest for fluid reasoning).
2. SH: One particular cultural group scores at least as high as the other group(s) on both the Afrikaans-culture and the Tswana-culture test version (differences between groups are smallest for short-term memory and attention, larger for working memory, and largest for fluid reasoning).

The urban Tswana children are exposed to aspects of both cultures and are therefore expected to score in between the Afrikaans and rural Tswana group on both test versions. This score pattern would be in line with both CCH and SH:

3. Urban Tswana children score in between Afrikaans and rural Tswana children on both the Afrikaans-culture and the Tswana-culture test version.

5. Method

5.1. 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 children from two primary schools in the town of Potchefstroom, North-West Province; 181 were black urban Tswana children from two primary schools in Ikageng, a township near Potchefstroom; 159 were black rural Tswana children from three primary schools in Ramatlabama, a rural setting 15 kilometres outside of the city of Mafikeng, North-West Province. The three groups differed in culture (Afrikaans-culture versus Tswana-culture), language (Afrikaans language versus Setswana language, the language of the Tswana), socioeconomic status (the Afrikaans group was of higher socioeconomic status than the two Tswana groups) and locality of residence (urban versus rural). The latter distinction, however, was not as large as the terms *urban* and *rural* might suggest; it was merely a relative distinction in that the children from Potchefstroom ("Afrikaans") and Ikageng ("urban Tswana") lived in a more urban area when compared to the more rural area of Ramatlabama ("rural Tswana"). However, we did not observe substantial differences in the quality of education unlike in many studies describing urban–rural distinctions. The main implication of our relative urban–rural distinction was a difference in the exposure to Afrikaans' culture; the urban Tswana group was more exposed to the Afrikaans way of life than the rural Tswana group.

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.

5.2. Instruments

Five cognitive tests were constructed: a short-term memory test, an attention test, a working memory test, a figural fluid

[1] The three groups included in the present study are characterized by large differences in SES and related environmental conditions, with the Afrikaans group living in the most advantaged environment. As a consequence, it is very likely that the Afrikaans group obtains higher overall scores than the other groups. Our hypotheses do not mention these conditions and resultant score differences, because we statistically correct for a presumably large part of these differences in our hypotheses tests by standardizing for SES.

Please cite this article as: Malda, M., et al., Rugby versus Soccer in South Africa: Content familiarity contributes to cross-cultural differences in cognitive test scores, *Intelligence* (2010), doi:10.1016/j.intell.2010.07.004

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-culture and a Tswana-culture 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 the Afrikaans and the Setswana language. Afrikaans children did all tests (so, both the tests with a more Afrikaans and the tests with a more Tswana content) in the Afrikaans language; Tswana children did both types of tests in the Setswana language. 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.

5.2.1. 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-culture and Tswana-culture test version used meaningful words with a higher familiarity for the Afrikaans and Tswana children, respectively. Examples of words used in the Afrikaans-culture test version are "computer", "camera" and "shower", and examples of the Tswana-culture test version are "tuckshop" (a small food shop that is common in the Tswana community), "soccer", and "braids".

5.2.2. 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-culture version targeted groups of two electrical appliances followed by one piece of clothing (e.g., Heater–Iron–Trousers); the Tswana-culture version aimed at groups of two family members followed by one animal (e.g., Aunt–Son–Dog).

5.2.3. Working memory test

This individually administered test is inspired by the Working Memory Test Battery for Children (Pickering & Gathercole, 2001). The first 3 items of our instrument 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 statements increased in number (from two to seven) as well as complexity. A statement of the lowest complexity consisted of one single sentence (e.g., A fridge is cold); a statement of a higher complexity consisted 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-culture and Tswana-culture 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-culture test version and "A soccer team has 11 players" for the Tswana-culture test version (soccer is the most popular sport among the Tswana children whereas rugby is most popular among the Afrikaans).

5.2.4. 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, 1998). 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-culture and Tswana-culture 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-culture test version, a drawing of a swimming pool was included, and the Tswana-culture 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.

5.2.5. 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-culture (e.g., rugby–swimming–cricket–tennis) and Tswana-culture 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.

5.2.6. Raven's Standard Progressive Matrices

The test (Raven et al., 1998) measures figural fluid reasoning and requires the child to complete a (meaningless)

Table 3
Means and standard deviations for scores on Raven's Standard Progressive Matrices.

Group	Children selected for			
	Afrikaans-culture version		Tswana-culture version	
	M	*SD*	*M*	*SD*
Afrikaans	23.75	5.07	23.75	4.83
Urban Tswana	14.75	5.31	13.00	4.78
Rural Tswana	12.71	4.94	12.90	4.99

Table 4
Means, standard deviations, and MANOVA for the raw cognitive scores.

Test	Afrikaans-culture version		Tswana-culture version		F (1, 499)	p
	M	SD	M	SD		
Short-term memory	8.36	2.04	8.59	2.03	1.59	.21
Attention	6.21	5.45	8.18	6.27	14.20	.00
Working memory	5.43	2.38	6.21	2.13	14.92	.00
Figural fluid reasoning	19.78	7.11	18.77	5.24	3.26	.07
Verbal fluid reasoning	11.70	2.89	12.21	3.34	3.45	.06

figural pattern and was collectively administered. The test was used as a reference point that did not reflect the Afrikaans-culture or Tswana-culture 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.

5.3. 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-culture and Tswana-culture), reflecting a 3×2 between subjects design (see Table 2). 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 (see Table 3 for means and standard deviations). 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 ($F[1, 159] = 0.00$, $p = 1.00$) and rural Tswana ($F[1, 157] = 0.06$, $p = .81$) 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-culture test version performed significantly higher than those selected for the Tswana-culture test version, $F(1, 179) = 5.41$, $p < .05$, partial $\eta^2 = .03$, Cohen's $d = .35$.

5.4. Perceived familiarity

Perceived familiarity was assessed as a manipulation check of the content 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-culture and Tswana-culture), 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.

5.5. Socioeconomic status

Children were asked six questions as an indication of 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.

5.6. Procedure

Eleven Afrikaans-speaking and eleven Setswana-speaking females were trained to administer the test battery. Seventeen

Table 5
Means and standard deviations for the raw cognitive scores.

Group	Short-term memory				Attention				Working memory				Figural fluid reasoning				Verbal fluid reasoning			
	Ac version		Tc version		Ac version		Tc version		Ac version		Tc version		Ac version		Tc version		Ac version		Tc version	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
A	8.71	1.59	7.95	1.75	7.53	6.45	9.75	6.55	6.98	2.41	6.53	2.01	26.49	3.06	20.09	3.89	13.88	2.15	13.93	2.78
UT	8.69	2.30	9.20	2.34	5.35	4.27	6.58	5.51	4.87	2.17	5.74	2.33	18.34	7.10	17.67	5.84	10.69	2.48	10.97	3.37
RT	7.66	1.95	8.56	1.71	5.87	5.39	8.40	6.41	4.55	1.79	6.42	1.92	14.83	4.53	18.68	5.47	10.68	2.79	11.87	3.12

Ac = Afrikaans-culture, Tc = Tswana-culture, A = Afrikaans, UT = Urban Tswana, RT = Rural Tswana.

Table 6
Means and standard deviations for the cognitive scores standardized for test version.

Group	Short-term memory				Attention				Working memory				Figural fluid reasoning				Verbal fluid reasoning			
	Ac version		Tc version		Ac version		Tc version		Ac version		Tc version		Ac version		Tc version		Ac version		Tc version	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
A	.17	.78	-.32	.86	.24	1.18	.25	1.05	.65	1.01	.15	.94	.94	.43	.25	.74	.75	.74	.51	.83
UT	.16	1.12	.30	1.15	-.16	.78	-.26	.88	-.23	.91	-.22	1.09	-.20	1.00	-.21	1.12	-.35	.86	-.37	1.01
RT	-.35	.96	-.02	.84	-.06	.99	.04	1.02	-.37	.75	.10	.90	-.70	.64	-.02	1.04	-.35	.97	-.10	.93

Ac = Afrikaans-culture, Tc = Tswana-culture, A = Afrikaans, UT = Urban Tswana, RT = Rural Tswana.

Table 7
Bivariate correlations of raw cognitive scores for the Afrikaans group.

	STM	ATT	WM	FigFR	VerbFr	Raven
STM	–	.36**	.31**	.37**	.34**	.31**
ATT	.08	–	.20	.26*	.38**	.22*
WM	.26*	.26*	–	.21	.25*	.31**
FigFR	-.04	.28*	.13	–	.26*	.31**
VerbFR	.09	.35**	.14	.34**	–	.43**
Raven	.07	.36**	.25*	.31**	.32**	–

Note. Values for the Afrikaans-culture test version are below the diagonal; values for the Tswana-culture test version are above the diagonal. Raven's Standard Progressive Matrices do not have an Afrikaans-culture and Tswana-culture test version. STM = Short-term memory, ATT = Attention, WM = Working memory, FigFR = Figural fluid reasoning, VerbFR = Verbal fluid reasoning, Raven = Raven's Standard Progressive Matrices. * $p<.05$. ** $p<.01$.

Table 8
Bivariate correlations of raw cognitive scores for the urban Tswana group.

	STM	ATT	WM	FigFR	VerbFr	Raven
STM	–	.65**	.49**	.64**	.07	.10
ATT	.42**	–	.62**	.52**	-.11	-.03
WM	.50**	.49**	–	67**	-.03	.02
FigFR	.52**	.57**	.57**	–	-.01	.11
VerbFR	.30**	.27*	.21*	.23*	–	.26*
Raven	.25*	.23*	.19	.46**	.34**	–

Note. Values for the Afrikaans-culture test version are below the diagonal; values for the Tswana-culture test version are above the diagonal. Raven's Standard Progressive Matrices do not have an Afrikaans-culture and Tswana-culture test version. STM = Short-term memory, ATT = Attention, WM = Working memory, FigFR = Figural fluid reasoning, VerbFR = Verbal fluid reasoning, Raven = Raven's Standard Progressive Matrices. * $p<.05$. ** $p<.01$.

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.

6. Results

Results are described in three sections. We first present preliminary analyses on item bias, score standardization, reliability, background variables, and perceived familiarity as manipulation check of content familiarity. This is followed by a validation of the cognitive structure that is suggested to underlie the test battery. Finally, a multivariate analysis of variance (MANOVA) is presented that tests the effects of group and test version on the cognitive test scores.

6.1. Preliminary analyses

6.1.1. Item bias

Item bias (differential item functioning) was computed in a logistic regression procedure in which item scores were predicted on the basis of score level, SES, group membership (dummy coded), and the interaction between score level and group membership. The analyses showed that 56 out of 212 items were significantly biased; however, the effect sizes were small; more specifically, the average effect size f^2 was .08 (range: .03–.30). The bias seemed to slightly favor the Afrikaans group; however, given the small effect sizes, we did not exclude any items from further analyses.

6.1.2. 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 the multivariate main effect of version was significant (Wilks' $\lambda = .92$, $F(5,495) = 9.15$, $p<.01$; see Table 4 for means, standard deviations, and MANOVA results); significantly higher scores were obtained for the Tswana-culture version of the attention test ($p<.01$, partial $\eta^2 = .03$, Cohen's $d = .34$) and the working memory test ($p<.01$, partial $\eta^2 = .03$, $d = .35$) than for the Afrikaans-culture 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. (Tables 5 and 6 show means and standard deviations for the unstandardized and standardized scores, respectively. Tables 7–9 display the correlations between the unstandardized test scores; correlations of these test scores with raw scores on Raven's Standard Progressive Matrices are also included.)

6.1.3. Reliability of cognitive tests

Table 10 gives an overview of the internal consistencies for all tests for each group and test version. Most values were acceptable to high. The highest values were found for the attention test, most probably due to the high score variation for

Table 9
Bivariate correlations of raw cognitive scores for the rural Tswana group.

	STM	ATT	WM	FigFR	VerbFr	Raven
STM	–	.18	.30**	.10	.14	.18
ATT	.36**	–	.40**	.33**	.16	.20
WM	.25*	.27*	–	.28*	.15	.16
FigFR	.01	.15	.08	–	.07	.41**
VerbFR	.14	.26*	.34**	.26*	–	.17
Raven	.05	.03	.04	.33**	.34**	–

Note. Values for the Afrikaans-culture test version are below the diagonal; values for the Tswana-culture test version are above the diagonal. Raven's Standard Progressive Matrices do not have an Afrikaans-culture and Tswana-culture test version. STM = Short-term memory, ATT = Attention, WM = Working memory, FigFR = Figural fluid reasoning, VerbFR = Verbal fluid reasoning, Raven = Raven's Standard Progressive Matrices. * $p<.05$. ** $p<.01$.

Table 10
Internal consistencies of all cognitive tests for each group.

Test and version	Overall group	Afrikaans	Urban Tswana	Rural Tswana	Δ[c]
Short-term memory[a]					
Afrikaans-culture	.75	.60	.79	.78	.30**
Tswana-culture	.74	.62	.81	.67	.32**
Attention[a]					
Afrikaans-culture	.94	.95	.94	.93	.16
Tswana-culture	.93	.95	.93	.91	.27**
Working memory[a]					
Afrikaans-culture	.78	.79	.78	.57	.33**
Tswana-culture	.75	.75	.77	.71	.11
Figural fluid reasoning[b]					
Afrikaans-culture	.89	.58	.89	.69	.61**
Tswana-culture	.77	.60	.81	.80	.34**
Verbal fluid reasoning[b]					
Afrikaans-culture	.61	.38	.46	.58	.18*
Tswana-culture	.70	.63	.71	.66	.11

[a]Values are split-half reliabilities corrected for test length with the Spearman-Brown formula. [b]Values are Cronbach's alpha values. [c]Δ gives the average effect size of the three comparisons of the differences in alpha values between the three groups (interpretable as Cohen's *d*); the asterisks refer to the significance of Hakstian and Whalen's (1976) *M* statistic that tests differences of the alpha values in the three groups. * $p<.05$. ** $p<.01$.

this test; the lowest values were found for the Afrikaans-culture test version of the verbal fluid reasoning test, showing less score variation. Significances between the reliabilities of the tests for the three cultural groups (Hakstian & Whalen, 1976) and their effect sizes (Liu & Weng, 2009) were computed. Seven out of ten comparisons showed a significant difference. The average absolute difference between any two coefficients was .10 (range: .01, .31); the average effect size (interpretable as Cohen's *d*) was .27 (range: .03 to .92), which points to a small effect size. It was concluded that although most tests yielded significant differences, the differences were not consequential.

6.1.4. 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 and verbal fluid reasoning (see Table 11). Girls scored higher on these tests than boys. A MANOVA with grade as independent variable showed significant main effects for all test scores (see Table 12). 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 $\eta^2=.50$ (post hoc analyses with Bonferroni criterion). Cohen's *d* was 2.14 for the comparison between the Afrikaans and urban Tswana group and 2.06 for the Afrikaans and rural Tswana group. SES correlated significantly with the sum score of each test (standardized for test version), except for short-term memory (see Table 13).

Table 11
Means, standard deviations, and MANOVA for sex.

Test	Female *M*	Female *SD*	Male *M*	Male *SD*	*F* (1, 499)	*p*	partial η^2	*d*
Short-term memory	.10	1.00	-.09	.99	4.56	.03	.01	.19
Attention	.05	1.05	-.05	.94	1.39	.24	.00	.00
Working memory	.07	1.07	-.07	.93	2.45	.12	.00	.00
Figural fluid reasoning	-.00	1.08	.00	.92	.00	.98	.00	.00
Verbal fluid reasoning	1.06	.98	-.10	1.00	5.49	.02	.01	.21

Table 12
Means, standard deviations, and MANOVA for grade.

Test	Grade 3 *M*	Grade 3 *SD*	Grade 4 *M*	Grade 4 *SD*	*F* (1, 499)	*p*	partial η^2	*d*
Short-term memory	-.18	1.01	.18	.96	16.29	.00	.03	.36
Attention	-.14	.99	.14	.99	10.50	.00	.02	.29
Working memory	-.19	.96	.20	1.00	19.92	.00	.04	.40
Figural fluid reasoning	-.17	1.01	.17	.96	14.44	.00	.03	.34
Verbal fluid reasoning	-.13	.97	.13	1.01	9.05	.00	.02	.27

6.1.5. Manipulation checks

Perceived familiarity was assessed as a manipulation check of the content familiarity of both test versions. The goal was 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-culture and Tswana-culture) 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-culture version was significantly higher than that of the Tswana-culture version, $F(1,\ 159)=35.89$, $p<.01$, partial $\eta^2=.18$, $d=.95$. 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-culture version, $F(1,\ 157)=9.71$, $p<.01$, partial $\eta^2=.06$, $d=.49$.

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-culture test version, $F(2,\ 250)=97.03$, $p<.01$, partial

Table 13
Bivariate correlations of standardized cognitive scores, socioeconomic status, and perceived familiarity.

Test	Socioeconomic status	Perceived familiarity
Short-term memory	.01	.17**
Attention	.16**	.24**
Working memory	.24**	.28**
Figural fluid reasoning	.32**	.40**
Verbal fluid reasoning	.35**	.29**

** $p<.01$.

Table 14
Means and standard deviations for the standardized cognitive scores after controlling for differences in sex, grade, and SES.

Group	Short-term memory				Attention				Working memory				Figural fluid reasoning				Verbal fluid reasoning			
	Ac version		Tc version		Ac version		Tc version		Ac version		Tc version		Ac version		Tc version		Ac version		Tc version	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
A	.17	.79	-.33	.87	.07	1.26	.09	1.05	.42	1.04	-.09	.99	.66	.55	-.06	.81	.42	.82	.17	.90
UT	.17	1.07	.30	1.16	-.06	.76	-.20	.90	-.10	.93	-.13	1.14	-.02	1.05	-.09	1.19	-.13	.94	-.25	1.08
RT	-.35	.94	-.02	.89	.02	.97	.10	1.01	-.27	.72	.20	.96	-.58	.71	.12	1.09	-.19	1.03	.03	1.04

Ac = Afrikaans-culture, Tc = Tswana-culture, A = Afrikaans, UT = Urban Tswana, RT = Rural Tswana.

$\eta^2 = .44$ (post hoc analyses with Bonferroni criterion). Cohen's *d* was 2.08 for the comparison between the Afrikaans and urban Tswana group and 1.91 for the Afrikaans and rural Tswana group. On the Tswana-culture test version, Afrikaans children scored significantly higher than the urban Tswana group, $F(2, 245) = 9.00$, $p < .01$, partial $\eta^2 = .07$ (post hoc analyses with Bonferroni criterion). Cohen's *d* was .33. 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-culture test version, the score differences between the Afrikaans group and the two other groups were smaller on the Tswana-culture version than on the Afrikaans-culture version, and the percentage of explained variance by familiarity was also substantially smaller. The manipulation checks largely supported the adequacy of familiarity differences of the test versions. Perceived familiarity scores (factor scores) correlated significantly with the standardized sum score of each test (see Table 13); as could be expected, higher familiarity was associated with higher test scores.

6.2. Validity of cognitive structure

Using structural equation modelling (Arbuckle, 2008), the validity of the suggested cognitive structure (Fig. 1) was tested. In order for the model to be identifiable, the regression weight of one of the indicators of the latent fluid reasoning factor was set to the value of one. For the overall sample ($N = 501$), with scores standardized for test version and for group, we found an excellent fit ($\chi^2[2, N = 501] = .48$, $p = .79$, $\chi^2/df = .24$, GFI = 1.00, AGFI = 1.00, TLI = 1.00, CFI = 1.00, RMSEA = .00); parameter estimates are given in Fig. 1. 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 an excellent fit when all parameters were identical ($\chi^2[17, N = 501] = 18.59$, $p = .35$, $\chi^2/df = 1.09$, GFI = .99, AGFI = .98, TLI = .99, CFI = 1.00, RMSEA = .01). Multigroup analyses testing the invariance of the model across the three groups showed that only the unconstrained model (configural invariance) provided an excellent fit ($\chi^2[6, N = 501] = 3.81$, $p = .70$, $\chi^2/df = .64$, GFI = 1.00, AGFI = .98, TLI = 1.00, CFI = 1.00, RMSEA = .00). The reason for the lack of fit of models with more invariance constraints is not clear.

6.3. MANOVA on cognitive test scores

To correct the cognitive test scores for the effects of sex, grade, and SES, we first performed linear regression analyses with these variables as independent variables (with dummy codes for sex and grade) and the scores on each test (standardized for test version) as dependent variables, and saved the standardized residual scores. Table 14 shows the means and standard deviations of these residual scores. In separate analyses, not further documented here, we determined whether group interacted with sex, grade, and SES for any of the tests. Negligible to small effect sizes were reported. This suggested that a correction for these variables did not distort the patterning of the group differences on which the next analysis focused.

A MANOVA was performed with test version (two levels: Afrikaans-culture and Tswana-culture) and group (three levels: Afrikaans, urban Tswana, and rural Tswana) as independent variables, and the residual scores as dependent variables (see Table 15). 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. Figs. 2–6 show the residual scores (i.e., mean *z* scores corrected for differences in sex, grade, and SES) and the significance of the score differences for each

Table 15
MANOVA on cognitive test scores.

Source	F^a	*p*	partial η^2
	Between subjects		
Group			
Short-term memory	8.61	.00	.03
Attention	2.35	.10	.01
Working memory	3.74	.02	.02
Figural fluid reasoning	13.55	.00	.05
Verbal fluid reasoning	11.30	.00	.04
Group x Test version			
Short-term memory	7.95	.00	.03
Attention	0.57	.57	.00
Working memory	10.03	.00	.04
Figural fluid reasoning	23.01	.00	.09
Verbal fluid reasoning	2.44	.09	.01
Error			
Short-term memory	(0.94)		
Attention	(0.99)		
Working memory	(0.95)		
Figural fluid reasoning	(0.87)		
Verbal fluid reasoning	(0.95)		

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).

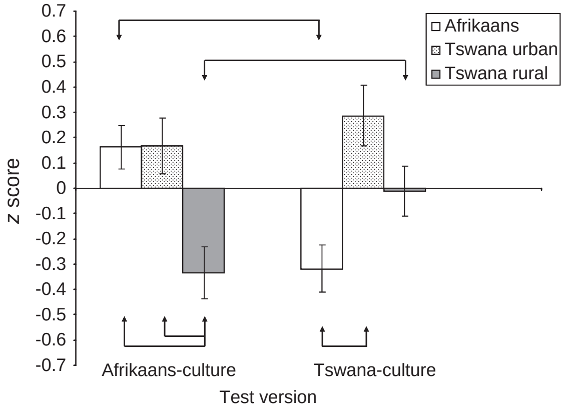

Fig. 2. Mean z scores for groups and test versions for short-term memory. *Note.* Arrows indicate significant differences (post hoc analyses with Bonferroni criterion). Significant differences between test versions: $F(1, 159) = 14.50$, $p < .01$ (Afrikaans group) and $F(1, 157) = 5.09$, $p < .05$ (rural Tswana group). Significant differences between groups: $F(2, 250) = 8.29$, $p < .01$ (Afrikaans-culture version) and $F(2, 245) = 8.42$, $p < .01$ (Tswana-culture version).

cognitive test in univariate tests. Table 16 provides an overview of the (dis)confirmation of hypothesis 1 to 3.

According to the first part of hypothesis 1, the Afrikaans children were supposed to score at least as high as the rural Tswana children on the Afrikaans-culture test version, with increasing score differences when going from left to right in Fig. 1. Figs. 2–6 show that the Afrikaans children scored significantly higher than the rural Tswana children on the Afrikaans-culture test version for each of the five tests, except for the attention test for which no significant differences were found. This means that hypothesis 1 was confirmed for all tests of the Afrikaans-culture test version, except for short-term memory. We observed significant differences for this test, which went in the expected direction (i.e., superior performance by Afrikaans children), whereas we expected relatively small or nonsignificant differences.

The second part of hypothesis 1 stated that the rural Tswana children were expected to score at least as high as the Afrikaans children on the Tswana-culture test version, with increasing score differences when going from left to right in Fig. 1. Figs. 2–6 show, however, that this group did not score significantly higher than the Afrikaans children on any of the tests of the Tswana-culture test version. This implies a disconfirmation of the second part of hypothesis 1 for two tests (working memory and fluid reasoning) for which we expected large differences. It seems that hypothesis 1 (CCH) was not completely supported.

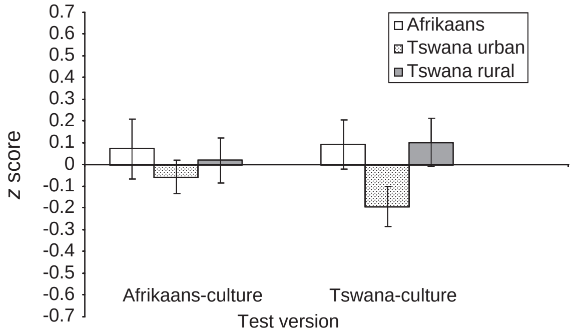

Fig. 3. Mean z scores for groups and test versions for attention.

Hypothesis 2 (SH) stated that one of the groups would score at least as high as the other groups on both test versions, with increasing score differences when going from left to right in Fig. 1. Fig. 2–6 show that, although the Afrikaans group scored at least as high as the rural Tswana group on the Afrikaans-culture version, this was not the case for the Tswana-culture test version. It seems that hypothesis 2 (SH) was not completely confirmed either.

Because neither CCH nor SH was fully supported, we decided to perform further analyses. In addition to our earlier examination of the patterning of group score differences for each of the two test versions, we addressed the patterning of scores on the test versions, as found within each of the three groups. CCH states that test content affects performance on cognitively (and with that, culturally) more complex tasks, with more familiar content leading to a higher performance than less familiar content. Therefore, the Afrikaans and rural Tswana groups are expected to do better on the version developed for their own group than on the version that was not specifically developed for their own group. This means that Afrikaans children would score at least as high on the Afrikaans-culture test version as on the Tswana-culture

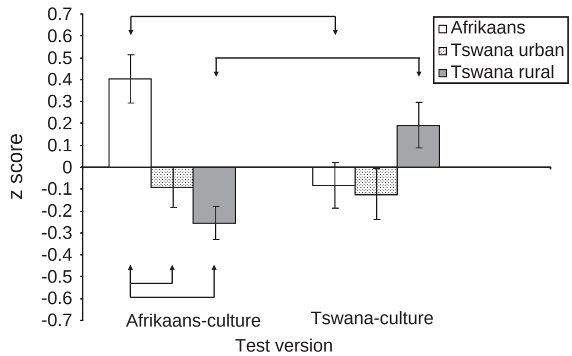

Fig. 4. Mean z scores for groups and test versions for working memory. *Note.* Arrows indicate significant differences (post hoc analyses with Bonferroni criterion). Significant differences between test versions: $F(1, 159) = 10.11$, $p < .01$ (Afrikaans group) and $F(1, 157) = 12.14$, $p < .01$ (rural Tswana group). Significant differences between groups: $F(2, 250) = 12.72$, $p < .01$ (Afrikaans-culture version).

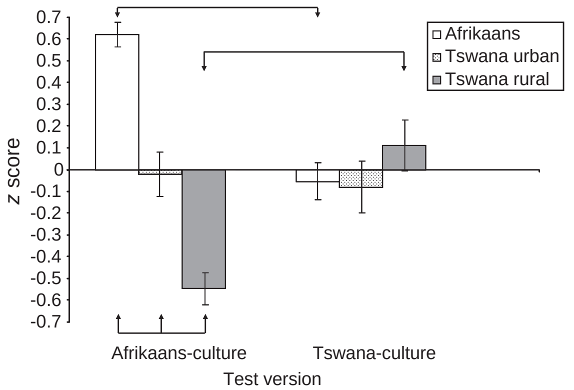

Fig. 5. Mean *z* scores for groups and test versions for figural fluid reasoning. *Note.* Arrows indicate significant differences (post hoc analyses with Bonferroni criterion). Significant differences between test versions: $F(1, 159) = 43.08$, $p < .01$ (Afrikaans group) and $F(1, 157) = 23.52$, $p < .01$ (rural Tswana group). Significant differences between groups: $F(2, 250) = 47.60$, $p < .01$ (Afrikaans-culture version).

version; rural Tswana children would score at least as high on the Tswana-culture version as on the Afrikaans-culture version. Differences between test versions are supposed to be smallest for short-term memory and attention, larger for working memory, and largest for fluid reasoning.

Fig. 2–6 show that 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. These findings were mostly in line with CCH; the expected score pattern (for comparison 1 in Table 16) was confirmed for all tests except for short-term memory (we expected very small differences) and for verbal fluid reasoning (we expected large differences). Although we expected very small score differences between the short-term memory versions, we want to stress that the direction of the differences was in line with CCH; Afrikaans children obtained higher scores on the Afrikaans-culture version than on the Tswana-culture version.

The rural Tswana children scored significantly higher on the Tswana-culture test version than on the Afrikaans-culture 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-culture and Tswana-culture version. These findings were mostly in line with CCH and confirmed our expectations for all tests (see the expected score pattern for comparison 2 in Table 16) except for short-term memory (we expected very small differences) and for verbal fluid reasoning (we expected large differences). Although we expected much smaller score differences between the short-term memory versions, we want to stress that the direction of the differences was in line with CCH; Tswana children obtained higher scores on the Tswana-culture version than on the Afrikaans-culture version.

The score patterns of the urban Tswana children mostly confirmed our expectations. They scored in between Afrikaans and rural Tswana children on the Afrikaans-culture test version of all tests (i.e., they did not score significantly higher than the Afrikaans children and lower than the rural Tswana children). Urban Tswana children did not score in between Afrikaans and rural Tswana children on the Tswana-culture test version of short-term memory and verbal fluid reasoning, however, for the other tests of the Tswana-culture version, there were no significant differences between the groups. Hypothesis 3 was confirmed for all tests except for short-term memory and verbal fluid reasoning of the Tswana-culture test version.

The urban Tswana children performed equally well on the Afrikaans-culture and Tswana-culture test version for all tests (confirming the expected score pattern for comparison 3 in Table 16), as could be expected based on their exposure to aspects of both Afrikaans and Tswana culture.

Tables 17 and 18 give an overview of Cohen's *d* values of the score differences between groups and the score differences between test versions, respectively. Score differences were expected to increase with cognitive complexity. When comparing the *d* values 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.

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-culture

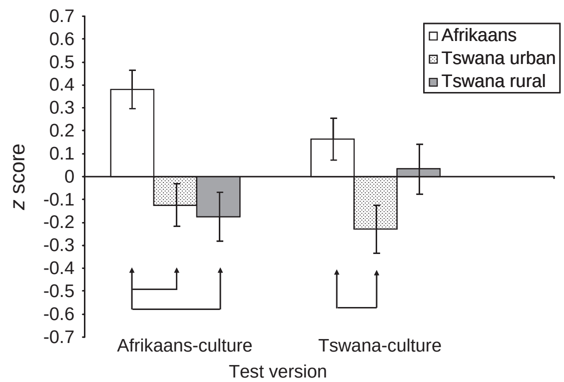

Fig. 6. Mean *z* scores for groups and test versions for verbal fluid reasoning. *Note.* Arrows indicate significant differences (post hoc analyses with Bonferroni criterion). Significant differences between groups: $F(2, 250) = 10.43$, $p < .01$ (Afrikaans-culture version) and $F(2, 245) = 3.85$, $p < .05$ (Tswana-culture version).

Please cite this article as: Malda, M., et al., Rugby versus Soccer in South Africa: Content familiarity contributes to cross-cultural differences in cognitive test scores, *Intelligence* (2010), doi:10.1016/j.intell.2010.07.004

Table 16
Level of confirmation of expected score patterns across test versions and groups.

Hypothesis	Test version	Expected score pattern	Cognitive test				
		Group	Short-term memory	Attention	Working memory	Figural fluid reasoning	Verbal fluid reasoning
1	Afrikaans-culture	Afrikaans ≥ Rural Tswana	no[a]	yes	yes	yes	yes
	Tswana-culture	Rural Tswana ≥ Afrikaans	yes	yes	no[b]	no[b]	no
2	Afrikaans-culture	One group ≥ other groups	no[a]	yes	yes	yes	yes
	Tswana-culture	One group ≥ other groups	no	yes	no	no	yes[c]
3	Afrikaans-culture	Urban Tswana in between other groups	yes	yes	yes	yes	yes
	Tswana-culture	Urban Tswana in between other groups	no	yes	yes	yes	no
Comparison	Group	Test version					
1	Afrikaans	Afrikaans-culture ≥ Tswana-culture	no[a]	yes	yes	yes	no[b]
2	Rural	Tswana-culture ≥ Afrikaans-culture	no[a]	yes	yes	yes	no[b]
3	Urban	Difference between versions smaller than for other groups	yes	yes	yes	yes	yes

Note. Hypothesis 1, 2, and 3, and comparison 1, 2, and 3 are correct for short-term memory and attention when score differences are smallest, for working memory when they are large(r) and for fluid reasoning when they are largest. [a]Score differences are in the expected direction but were not expected to be significant for short-term memory. [b]Score differences were not significant. [c]Only for the Afrikaans and urban Tswana group.

version of figural fluid reasoning ($F[1, 160] = 155.26$, $p < .01$, partial $\eta^2 = .49$, Cohen's $d = 1.96$). Figural fluid reasoning showed the largest differences between test versions in both the Afrikaans group ($F[1, 159] = 43.08$, $p < .01$, partial $\eta^2 = .21$, $d = 1.04$) and the rural Tswana group ($F[1, 157] = 23.52$, $p < .01$, partial $\eta^2 = .13$, $d = -.76$). It appears, in summary, that the confirmation of CCH and disconfirmation of SH is strongest for the cognitively most complex tests, because the familiarity effects are largest for these tests.

7. 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-culture and a Tswana-culture 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 children generally scored higher on the Afrikaans-culture version than the Tswana children. Tswana children however, did not significantly score higher than the Afrikaans children on the Tswana-culture test version. Nevertheless, most performance differences between the groups were smaller on this version than on the Afrikaans-culture version. 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-culture and Tswana-culture test version, showing that these children have enough knowledge of both cultures to perform equally on both versions. We can conclude that our results support the idea that (after correcting for confounding differences in SES) 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 on the basis of the test's low cognitive complexity. Urban Tswana children scored highest on both versions of this test. We could not capture any educational characteristics to 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

Table 17
Cohen's d for cognitive score differences between cultural groups.

Test	Afrikaans-culture test version			Tswana-culture test version		
	A - RT	A - UT	UT - RT	A - RT	A - UT	UT - RT
Short-term memory	.59	-.00	.51	-.35	-.61	.30
Attention	.05	.13	-.09	-.01	.30	-.31
Working memory	.77	.53	.20	-.29	.04	-.32
Figural fluid reasoning	1.96	.81	.62	-.18	.03	-.18
Verbal fluid reasoning	.65	.62	.06	.14	.42	-.26

A = Afrikaans, RT = Rural Tswana, UT = Urban Tswana.

Table 18
Cohen's d for cognitive score differences between test versions.

Test	Afrikaans group	Urban Tswana group	Rural Tswana group
Short-term memory	.60	-.11	-.36
Attention	-.02	.17	-.09
Working memory	.50	.03	-.55
Figural fluid reasoning	1.04	.06	-.76
Verbal fluid reasoning	.28	.11	-.22

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 closer examination, there were significant score differences on the Afrikaans-culture version and no significant differences on the Tswana-culture version (interactions were disordinal). One could argue that the difficulty level of the Tswana version was lower than that of the Afrikaans-culture 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-culture version than on the Afrikaans-culture version. Therefore, we conclude that our findings are not in line with SH, and that (after correcting for confounding differences in SES) it is cultural rather than cognitive complexity that explains more differences between groups, providing support for CCH.

Our study fits in a series 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 a high 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 three limitations. First, the between-subjects design limits the direct comparability of the two test versions. Second, the results show that the Afrikaans-culture and Tswana-culture 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. Third, 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 Fig. 1), probably more tests would need to be included. A more extensive battery of tests would enable stronger conclusions on the validity of SH.

Our study has some practical implications that should be taken into account in cross-cultural assessment in general, but more specifically in assessment in multicultural societies where it is common to derive (cognitive) tests from the cultural background of a single group, usually the majority group. 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, test selection should be based on the appropriateness of content whenever possible. The content familiarity of a cognitive test should be taken into account in evaluations of a child's performance and in cross-cultural comparisons of 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.

References

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

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

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

Bridgeman, B., & Buttram, J. (1975). Race differences on nonverbal analogy test performance as a function of verbal strategy training. *Journal of Educational Psychology*, *67*, 586−590.

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.

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.

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.

Fagan, J. F., & Holland, C. R. (2002). Equal opportunity and racial differences in IQ. *Intelligence*, *30*, 361−387.

Please cite this article as: Malda, M., et al., Rugby versus Soccer in South Africa: Content familiarity contributes to cross-cultural differences in cognitive test scores, *Intelligence* (2010), doi:10.1016/j.intell.2010.07.004

Fagan, J. F., & Holland, C. R. (2007). Racial equality in intelligence: Predictions from a theory of intelligence as processing. *Intelligence*, *35*, 319−334.

Fagan, J. F., & Holland, C. R. (2009). Culture-fair prediction of academic achievement. *Intelligence*, *37*, 62−67.

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.

Hakstian, A. R., & Whalen, T. E. (1976). A k-sample significance test for independent alpha coefficients. *Psychometrika*, *41*, 219−231.

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.

Hunt, E., & Carlson, J. (2007). Considerations relating to the study of group differences in intelligence. *Perspectives on Psychological Science*, *2*, 194−213.

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.

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. (1993). Spearman's hypothesis tested with chronometric information-processing tasks. *Intelligence*, *17*, 47−77.

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

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

Liu, H. Y., & Weng, L. J. (2009). An effect size index for comparing two independent alpha coefficients. *British Journal of Mathematical and Statistical Psychology*, *62*, 385−400.

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.

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.

Montie, J. E., & Fagan, J. F. (1988). Racial differences in IQ: Item analysis of the Stanford-Binet at 3 years. *Intelligence*, *12*, 315−332.

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.

Pickering, S. J., & Gathercole, S. E. (2001). *Working Memory Test Battery for Children.* London: Pearson Assessment.

Raven, J., Raven, J. C., & Court, J. H. (1998). *The Standard Progressive Matrices*. San Antonio, TX: Psychological Corporation.

Rushton, J. P. (2002). Jensen effects and African/Coloured/Indian/White differences on Raven's standard progressive matrices in South Africa. *Personality and Individual Differences*, *33*, 1279−1284.

Rushton, J. P., Bons, T. A., Vernon, P. A., & Cvorovic, J. (2007). Genetic and environmental contributions to population group differences on the Raven's progressive matrices estimated from twins reared together and apart. *Proceedings of the Royal Society: B*, *274*, 1773−1777.

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

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

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.

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.

Te Nijenhuis, J., Evers, A., & Mur, J. P. (2000). Validity of the differential aptitude test for the assessment of immigrant children. *Educational Psychology*, *20*, 99−115.

Te Nijenhuis, J., Tolboom, E., Resing, W., & Bleichrodt, N. (2004). Does cultural background influence the intellectual performance of children from immigrant groups? *European Journal of Psychological Assessment*, *20*, 10−26.

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.

Te Nijenhuis, J., & Van der Flier, H. (2003). Immigrant-majority group differences in cognitive performance: Jensen effects, cultural effects, or both? *Intelligence*, *31*, 443−459.

Te Nijenhuis, J., & Van der Flier, H. (2004). The use of safety suitability tests for the assessment of immigrant and majority group job applicants. *International Journal of Selection and Assessment*, *12*, 230−242.

Te Nijenhuis, J., & Van der Flier, H. (2005). Immigrant-majority group differences on work-related measures: The case for cognitive complexity. *Personality and Individual Differences*, *38*, 1213−1221.

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.

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

Wicherts, J. M., & Dolan, C. V. (2010). Measurement invariance in confirmatory factor analysis: An illustration using IQ test performance of minorities. *Educational Measurement: Issues & Practice*, *29, 3*, 39−47.

Wicherts, J. M., Dolan, C. V., Carlson, J. S., & Van der Maas, H. L. J. (2010). Raven's test performance of sub-Saharan Africans: Average performance, psychometric properties, and the Flynn Effect. *Learning and Individual Differences*, *20*, 135−151.

Wicherts, J. M., & Johnson, W. (2009). Group differences in the heritability of items and test scores. *Proceedings of the Royal Society: B*, *276*, 2675−2683.

Please cite this article as: Malda, M., et al., Rugby versus Soccer in South Africa: Content familiarity contributes to cross-cultural differences in cognitive test scores, *Intelligence* (2010), doi:10.1016/j.intell.2010.07.004