Tilburg University

Meta-analysis of cross-cultural comparisons of cognitive test performance

van de Vijver, F.J.R.

Published in:
Journal of Cross-Cultural Psychology

Publication date:
1997

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
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.

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.

A meta-analysis of cross-cultural comparisons of cognitive test scores published between 1973 and 1994 is presented. Larger performance differences were found in cross-national than in intranational (cross-ethnic) comparisons. There was a strong age bias, with an overrepresentation of primary-school-age children. Five models to explain cross-cultural differences in cognitive performance were examined. First, these differences showed a positive relationship with differences in affluence of the cultural groups. Second, performance differences increased with chronological and educational age. Third, only in intranational studies were performance differences positively related to the complexity of stimuli. Fourth, performance differences were larger on common Western tasks than on locally developed non-Western tasks. Finally, performance differences were not larger for abstract thinking than for other cognitive domains. Cross-cultural performance differences were better predicted by task characteristics in intranational comparisons and by population characteristics in cross-national comparisons.

META-ANALYSIS OF CROSS-CULTURAL COMPARISONS OF COGNITIVE TEST PERFORMANCE

FONS VAN DE VIJVER
Tilburg University

In the first 25 volumes of the *Journal of Cross-Cultural Psychology*, 6.4% of the publications describe cognitive studies (Öngel & Smith, 1994); this figure has been fairly stable across the 25 years. Even though this may seem a modest number, cognition ranks fourth after social psychology (44.2%), personality (12.5%), and methodology (12.5%). Such a pattern of a modest though stable interest in cross-cultural comparisons in the cognitive area will probably also be found in other publications.

Comparative studies using mental tests have produced a wealth of data. However, as often observed in psychology, studies of the same topic do not always show similar or even convergent findings. Meta-analysis has been proposed as a useful tool to integrate research findings (e.g., Hedges & Olkin, 1985). This article presents the results of a meta-analysis of cross-cultural comparisons of test scores on cognitive instruments published between 1973

AUTHOR'S NOTE: Address correspondence to Fons Van de Vijver, Department of Social Sciences, P.O. Box 90153, 5000 LE Tilburg, the Netherlands; phone: +31-13-466 2528; fax: +31-13-466 2370; e-mail: fons.vandevijver@kub.nl.

JOURNAL OF CROSS-CULTURAL PSYCHOLOGY, Vol. 28 No. 6, November 1997 678-709
© 1997 Western Washington University

and 1994. The appropriateness of explanatory models of cross-cultural differences in cognitive test scores is examined.

At least five models have been proposed to account for cross-cultural differences in cognitive test performance. The first can be labeled the *universalistic model* (cf. Berry, Poortinga, Segall, & Dasen, 1992). In this model, the basic properties of the human cognitive system are universal and cultures shape these universal properties in their own way. Intergroup differences in scores on mental tests reflect this differential practice. Factor-analytic studies of intellectual tasks provide an example. In a review of factor-analytic studies of cognitive test batteries, Irvine (1979) showed that there is a stable, cross-culturally invariant structure underlying the often massive differences in means across groups.

Segall, Dasen, Berry, and Poortinga (1990) coined the term *great divide model* to denote models that question the universality of psychological processes. When applied to cognitive test performance, the model can be taken to refute the universality of cognitive processes. Tulviste (1991), drawing on work by Luria and Vygotsky, argues that abstract, logical thinking is not a universal attainment but is acquired through literacy and education. According to Buck-Morss (1975), formal operational thinking is restricted to Western, industrialized countries: "Abstract, formal recognition may reflect a particular social structure, embodying the principles of exchange value, reification, and alienation that govern production and exchange in the industrialized West" (p. 35). In a similar vein, Chapman (1988) has criticized the assumed universalism of formal operations. Cognitive development cannot be construed as a unidirectional, teleological progress toward a universal end state (i.e., formal operational thinking). Ontogenetic development gradually moves away from a universal initial state (at birth); the direction and outcome of this development vary across cultural contexts.

The *cognitive complexity model* is based on Spearman's (1923) work. He hypothesized that tasks with a higher level of cognitive complexity reveal larger cross-cultural performance differences. The cognitive complexity of a task can be operationalized as its loading on the first factor in a factor analysis of scores on a cognitive test battery; more complex tasks have higher loadings. Cognitive complexity is not the same as task difficulty; the latter is usually measured by the percentage of subjects successfully carrying out a task (cf. Jensen, 1980). The cognitive complexity of solving the analogy black:white::night:_ is higher than that of reacting to a visual stimulus within 50 milliseconds, although fewer individuals may successfully complete the latter task. In the former, more complex stimulus transformations (and hence more abstract thinking) are required to arrive at a solution than in the latter. Tasks requiring more abstract thinking have a higher cognitive complexity.

The model is somewhat related to the great divide model. The major difference is the nature of the cross-cultural differences to be expected: In the area of abstract thinking, the great divide model predicts qualitative differences across cultures and the cognitive complexity model predicts quantitative differences. Furthermore, the cognitive complexity model primarily involves the patterning of cross-cultural differences across a variety of tasks (larger differences for more complex tasks), whereas the great divide model focuses more on specific cognitive attainments such as abstract thinking.

In the 1970s, Jensen (e.g., 1977) proposed the so-called *cumulative deficit model*. Performance differences between children of low and high socioeconomic status are hypothesized to increase with age. Analogously, a *cumulative differences model* can be proposed. The model holds that cross-cultural differences in cognitive test performance increase with age.

The previous models implicitly assume that cross-cultural differences in test scores are valid indicators of intergroup differences in underlying competence. The *bias model* questions the accuracy of the assumption. This model proposes that cross-cultural score differences are due to inadequacies of measuring instruments. An example can be found in Jensen (1980); he emphasizes the need to choose appropriate stimulus materials. It is unfortunate that approaches that focus on the choice of stimulus material often do not specify testable conditions to check the accuracy of the stimulus material. Another example of the bias model can be found in Mercer (1984). Her work can be seen as a reaction to the numerous cross-cultural score comparisons that are reported in the literature in which the appropriateness of the instruments were not examined. She claims that strict criteria have to be met before test scores can be compared across cultures. The test should have identical psychometric characteristics across all cultural groups involved. Among other things, this amounts to the cross-cultural invariance of the reliability, factor pattern and structure, temporal stability (test-retest correlation), patterning of item difficulty (i.e., the absence of any item by cultural group interaction in an analysis of variance), and validity. The appropriateness of the criteria can be questioned. Ideally, all these criteria have to be met. From a practical perspective, however, it is unlikely that any common psychological test can live up to these criteria, even in applications in culturally related groups. Examples of less restrictive bias models can be found in the literature on differential item functioning (e.g., Berk, 1982; Camilli & Shepard, 1994; Holland & Wainer, 1993).

Tests of the accuracy of the five models are hampered by at least two factors. First, the models have complex relationships. The universalistic and great divide models provide competing explanations, whereas the other models are compatible with each other and the universalistic and great divide

models. For example, performance differences may be due to item bias on some items and to another model on other items. Another problem is a lack of clarity that plagues all models. For example, does the great divide model state that certain cognitive skills are not acquired in certain cultures, that these skills are acquired by some in particular cultures and by all in other cultures, or that the skills are acquired by all members of all cultures though at different ages and levels of mastery? Due to these problems, a meta-analysis cannot provide a Popperian critical test of the accuracy of the models and can merely examine the accuracy of corollaries of the models.

The universalistic model claims that basic cognitive processes are universal and that score differences of groups are brought about by differential cultural practices, notably educational differences. In the framework of a meta-analysis, cross-cultural differences in Gross National Product (GNP) and educational expenditure (i.e., amount of money spent on education per capita per year) can serve as proxies for educational variables (cf. Ballantine, 1993, p. 329). This leads to the following hypothesis:

Hypothesis 1: Culture-level variables such as differences in GNP and educational expenditure (per capita) between cultural groups are positively related to differences on cognitive test scores.

Great divide models have focused on cross-cultural comparisons in formal-operational thinking of groups with pronounced differences in educational background such as comparisons of literate and nonliterate subjects. Therefore, the following corollary of the model is tested here:

Hypothesis 2: Cross-cultural performance differences of literate and nonliterate individuals are larger on tests of abstract thinking than on other cognitive tests.

Former studies of the cognitive complexity model have examined task complexity. However, the latter can be taken to confound complexity levels of stimuli and responses (e.g., the response format used). The contributions of both are examined separately in the present study:

Hypothesis 3a: Cross-cultural performance differences are positively related to stimulus complexity.

Hypothesis 3b: Cross-cultural performance differences are positively related to response complexity.

Age shows a strong positive relationship (at least up to adulthood) with cross-cultural performance differences according to the cumulative differ-

ences model. Because it may be impossible to disentangle chronological and educational age, a hypothesis of their combined effects is tested:

Hypothesis 4: Cross-cultural performance differences are positively related to age and years of schooling.

The bias model scrutinizes the appropriateness of instruments in a cross-cultural context. One of the leading themes in the literature on bias has been the detection of explicit and implicit references in an instrument to the (typically Western) cultural background of test composers. These references are taken to jeopardize the suitability of the instrument elsewhere. The adequacy of this reasoning is tested in the last hypothesis:

Hypothesis 5: Cross-cultural performance differences are larger for common Western tasks than for locally developed, non-Western tasks.

METHOD

LITERATURE REVIEW

The period covered in the meta-analysis ranges from 1973 to 1994. Studies were gathered from several sources. Most were retrieved from *PsycLit*, using *cross-cultural differences* and *intel** or *cognit** as keywords. In addition, the *South African Journal of Psychology*, relevant handbooks (e.g., Irvine & Berry, 1988), and Ph.D. theses (e.g., Van der Flier, 1980) were searched. Studies were included if data were available about the instrument(s) administered and the averages and standard deviations in at least two cultural groups. A *t* statistic or *F* ratio of a test of equal average scores of the two groups or the biserial correlation between test performance and cultural group were used (if reported) when studies involved two groups and no averages and standard deviations were provided.

Most cross-cultural cognitive studies apply more than one test, subtest, or instrument. For the present purposes, such studies were split into pairwise comparisons. The selection was made in such a way that scores of a group on a particular instrument were used only once. When an odd number of cultural groups was involved, the data of a randomly selected group on the instrument were omitted. For instance, when five groups were involved, a randomly chosen group was not considered; the remaining four samples were then split into two randomly chosen pairwise comparisons. The randomiza-

tion process was repeated for each instrument reported. The selection and randomization procedure ensured independence of the analyzed data.

When both subtest and total test scores were given in a study (which is frequently the case in examinations applying intelligence tests), only the former were considered so as to maximize the size of the database while avoiding dependencies among the data. An attempt was made to eliminate overlapping data that appeared in more than one publication.

There were almost no studies for which all data were available. Of the several hundreds of relevant studies, only 197 contained at least the minimal data. On the basis of these studies, 1,555 independent comparisons were composed.

MEASURES

The following measures were derived from the original publication: reference to the source, year of publication, average scores and standard deviations of the cognitive test, the average chronological and educational ages, gender composition of the sample (males, females, or both), and type of publication (book or journal).

Each instrument applied was rated on three dimensions. The first one referred to the type of task. Three types were distinguished: common Western tasks (including commercially available or widely applied Western tests such as the Raven and Wechsler tests), locally developed Western tasks, and locally developed non-Western tasks.

Second, the cognitive complexity of stimuli was rated. The ratings were based on reported loadings in large factor analytic studies (e.g., Carroll, 1993; Jensen, 1980; Vernon, 1950, 1979). Some arbitrariness in the rating of tests could not be avoided because, for obvious reasons, there is no single study in which all tasks reported in the literature have been administered. The stimulus complexity was scored as follows (in order of increasing complexity):

1. simple reconstruction, simple reaction times, substitution, and speed-and-accuracy tasks (examples are the Coding subscale of the Wechsler Intelligence Scale for Children-Revised [WISC-R] and the Digit Symbol subscale of the Wechsler Adult Intelligence Scale [WAIS]);
2. questions of factual knowledge, simple memory (e.g., digit span forward), drawing, simple arithmetic, and free recall (examples are paired associate tasks and WISC-R Vocabulary and WAIS Information);
3. embedded problems, block patterns, mazes, complex memory (digit span backward), "closure," and jigsaw puzzles (examples are Block Designs and Object Assembly of the WISC-R);

4. sorting, classification, analysis, and spatial orientation (examples are various sorting tasks and WISC-R Picture Arrangement);
5. insight, understanding, series extrapolations, analogies, and reasoning (examples are the Raven tests and Cattell's Culture Fair Intelligence Test).

Third, the response complexity was rated as follows (in order of increasing complexity):

1. multiple choice, simple manual operations (e.g., pressing a button), and simple writing (examples are reaction time tasks and the Raven tests);
2. open end with simple answer (e.g., single words), and complex manual operations (examples are most of the subtests of the Wechsler Memory Scales);
3. free description (examples are WISC-R Vocabulary and WAIS Comprehension).

The interrater reliability of the latter two coding schemes was evaluated on a set of 29 instruments of the database. The stimulus complexity scores of two independent raters showed an intraclass correlation of .93 (66% agreement of scoring category); the value for response complexity was .97 (93% agreement). These values were deemed adequate.

The following measures were derived from other sources: latitude of the countries' capital, GNP and expenditure on formal education (per capita) in 1986¹ (United Nations, 1988), and scores on individualism-collectivism, power distance, uncertainty avoidance, and masculinity-femininity as reported by Hofstede (1980). The latter were chosen as one of the very few databases with psychological data for many countries.

A distinction was made between intranational and cross-national studies, depending on whether individuals from one or more nations were included in the study. Intranational comparisons were more numerous (intranational $n = 1,210$; cross-national $n = 345$).

Estimates of effect sizes were usually based on means and standard deviations. Both in intranational and cross-national studies, an effect size was defined as the absolute difference between averages obtained by two cultural groups, M_1 and M_2 , divided by the pooled standard deviation s_p : $d = |M_1 - M_2| / s_p$, in which $s_p = \sqrt{\{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2\} / (n_1 + n_2 - 2)}$ (n_1 and n_2 are the sample sizes of the two groups and s_1 and s_2 are their standard deviations). To deal with the problem that occurs particularly in small samples, the observed effect size overestimates its population counterpart, a bias correction was applied: $d^* = d \{1 - \frac{3}{4(n_1 + n_2) - 9}\}$ (Hedges & Olkin, 1985).

RESULTS

Analyses of descriptive measures, including study characteristics are presented first, followed by a test of the hypotheses.

DESCRIPTIVE MEASURES

It was found that 81% of the publications had appeared in journals and 19% in books. A *t* test was carried out to compare the effect sizes of books ($n = 307$) and journals ($n = 1,248$). The averages were 0.59 ($SD = 0.51$) and 0.65 ($SD = 0.62$), respectively; the corresponding *t* value was not significant, $t(1,553) = -1.63$, *ns*. It was decided to combine data of both sources of publications.

The main study characteristics have been presented in Figures 1, 2, and 3. As can be seen in Figure 1a, the number of independent comparisons showed a high variability and a positive skewness for both cross-national and intranational studies. The averages of the cross-national and intranational means were 6.57 and 8.60, respectively; these did not differ significantly, $t(195) = .70$, *ns* (equality of means of intranational and cross-national studies was tested for 6 study characteristics; per test, an error rate of .05/6 was used to obtain a family-wise error rate of .05). The mode of the distribution was two for both types of studies. If studies with more than 16 comparisons are seen as a single type of study, the distribution is bimodal. This observation points to two types of studies: one comparing cultural groups on one or two target variables and one comparing cultural groups on a wide variety of measures (frequently omnibus intelligence tests, as documented later on).

A weighted procedure had to be applied to determine the average sample size in order to deal with the problem that pairwise comparisons of a study involving more than two cultural groups have different sample sizes for each group. The weights applied were proportional to the number of comparisons of the study; for example, when eight independent comparisons were derived from a study, each comparison was given a weight of 1/8. The sample sizes also showed a bimodal distribution, with one mode for studies of up to 50 subjects and another for studies of more than 500 subjects (Figure 1b). The averages of the intranational studies (302.00) and cross-national studies (283.32) did not differ significantly, $t(387) = .32$, *ns*.

The age distributions of cross-national and intranational studies have been presented in Figure 1c. There is a strong overrepresentation of primary-school-age children in cross-cultural cognitive research; the modes of both distributions was 10 years. A *t* test of the difference in mean age of subjects

in these studies did not show a significant result, $t(174) = -1.71$, *ns*. It appears that cross-cultural cognitive research has implicitly focused on primary-school-age children. The strong age bias could be engendered by at least two factors. First, theoretical frameworks of cross-cultural cognitive differences often have a developmental focus such as Piagetian theory. The bias could also be due to convenience sampling (Van de Vijver & Leung, 1997). The recruitment of subjects is then more determined by efficiency and availability than by theory; convenience sampling often amounts to the choice of schools to recruit subjects as schools enable an efficient way of data collection among large groups of subjects.

Country characteristics have been presented in Figure 2. The countries most frequently involved in cross-cultural comparisons are the United States (29%), the United Kingdom (11%), Japan (8%), Canada (5%), and China (4%) (see Figure 2a). The high rankings of the United States, Japan, and China are mainly due to the huge number of studies comparing U.S. groups with Japanese and Chinese groups. The predominance of the United States was also observed in intranational studies (see Figure 2b). Other plural societies with relatively much intranational research are Australia, Israel, South Africa, and the United Kingdom.

The distribution of the GNPs is presented in Figure 2c. Two curves have been drawn. The first gives the frequency distribution for the 48 countries represented in the data; the curve is called unweighted in the figure as each country provides a single entry. However, as is clear from Figure 2a and 2b, not all 48 countries are equally often represented in the database. Therefore, a second curve was based on the 1,555 comparisons of the database. Each comparison yielded two GNPs, making a total of 3,110 GNPs; their distribution, called the weighted curve, is also presented in Figure 2c. The difference between the curves is striking. The unweighted curve shows an overrepresentation of the very low and very high GNPs, whereas the weighted curve shows a much stronger overrepresentation of affluent countries. More precisely, 15% of the countries in the database have a GNP of more than (U.S.) \$11,000 but they constitute 65% of the countries involved in cross-cultural score comparisons. This points to a strong bias toward affluent countries.

Task characteristics have been presented in Figure 3. The most frequently employed tests in cross-national studies were (in decreasing order of importance) the WISC (original and revised), Raven tests, Kaufman Assessment Battery for Children, Stanford-Binet Intelligence Scale, and Wechsler Adult Intelligence Test (original and revised) (see Figure 3a). The order for the intranational studies was the WAIS (original and revised), the WISC (original and revised), Raven tests, McCarthy Scales of Children's Abilities, and Columbia Mental Maturity Scale (see Figure 3b). In sum, intelligence tests,

in most cases omnibus tests, are the most commonly employed instruments in cross-cultural research of mental abilities.

An overview of the stimulus complexity of tasks is presented in Figure 3c. The distribution of stimulus complexity was not uniform. The first and fourth levels were relatively underrepresented, whereas the second and fifth levels were overrepresented. The overrepresentation is mainly a consequence of the frequent cross-cultural applications of general intelligence tests, which usually consist of subtests of the second (e.g., vocabulary and information subtests) and fifth (e.g., series extrapolations and analogies) levels of stimulus complexity. In addition, most memory tasks that were applied pertained to the second complexity level (e.g., digit span).

As can be seen in Figure 3d, response complexity showed a strong, negative correlation with frequency of application; more complex response formats were less frequently applied than multiple choice, short, oral, or written answers.

The curves of the cross-national and intranational studies intersect in both Figures 3c and 3d, possibly pointing to an interaction. The significance of the interaction was examined in two loglinear analyses. In the first one, stimulus complexity (5 levels) and intranational versus cross-national comparisons (2 levels) were the independent variables and frequency was the dependent variable. Removal of the interaction component from the saturated model led to a poor fit, Pearson $\chi^2(4) = 48.41, p < .01$, indicating that the curves of the intranational and cross-national studies cannot be taken to run parallel. The second analysis, using the same independent variables, examined response complexity. Again, a poor fit of a model without interaction component was observed, Pearson $\chi^2(2) = 24.23, p < .01$. It appears that the interactions of both analyses were significant. As can be seen in Figures 3c and 3d, cross-national comparisons were more often based on easy tasks, whereas more complex tasks were relatively more often applied in intranational comparisons. Easier answer formats were more common in cross-national comparisons, whereas more complex answer formats were relatively more common in intranational comparisons.

Both males and females participated in 86% of the studies, 5% involved only females and 9% only males. It is not unlikely that overall, more males than females were involved. However, it is difficult to estimate relative proportions because many studies do not specify the numbers per gender.

Effect sizes are presented in Figure 4. The distribution showed a strong, positive skewness of 2.98 ($z = 47.88, p < .01$). The mode was 0 and the median effect size was 0.51. If Cohen's (1962) definition of small, medium, and large effects is applied (with cutoff points of 0.35 and 0.65, respectively), each category contained about one third of the studies.

The equality of effect sizes in intranational and cross-national studies was compared in a t test. A significant t value of 4.19 ($df = 1553$, $p < .01$) was observed. The average effect size was 0.58 for intranational studies and 0.71 for cross-national studies; intergroup performance differences were significantly smaller in intranational than in cross-national studies. The difference can also be seen in Figure 4. The curve of the intranational studies is higher for small effect sizes and lower for large effect sizes than the curve of cross-national comparisons. Born, Bleichrodt, and Van der Flier (1987), studying gender differences in cognition, also found that intranational studies tended to produce smaller differences. Moreover, effect sizes of intranational comparisons were significantly more homogeneous than those of cross-national comparisons, intranational $SD = 0.49$, cross-national $SD = 0.64$; $F(344, 1,209) = 1.66$, $p < .01$.

Finally, historical shifts in praxis were examined. Two multiple regression analyses were carried out, one for intranational and one for cross-national comparisons. Publication year was the dependent variable and type of task, stimulus complexity, response complexity, average age of subjects, sample size, and number of tasks administered were independent variables. Both analyses showed a significant multiple correlation, intranational $R^2 = .15$, $F(6, 640) = 18.51$, $p < .01$, and cross-national $R^2 = .22$, $F(6, 297) = 13.63$, $p < .01$. As can be seen in Table 1, more recent intranational studies applied more tasks, $\beta = 0.23$, $t(640) = 5.97$, $p < .01$; of local origin, $\beta = 0.28$, $t(640) = 7.20$, $p < .01$; with lower response complexities, $\beta = -0.12$, $t(640) = -3.06$, $p < .01$; to older subjects, $\beta = 0.18$, $t(640) = 4.88$, $p < .01$. The tendency to apply more instruments was also observed in cross-national studies, $\beta = 0.25$, $t(297) = 4.63$, $p < .01$. Additionally, these studies showed a tendency to apply tasks with less complex stimuli, $\beta = -0.12$, $t(297) = -2.26$, $p < .01$; to larger samples, $\beta = 0.36$, $t(297) = 6.63$, $p < .01$. A test of the equality of the regression lines of intranational and cross-national studies showed a highly significant value, $F(7, 937) = 10.30$, $p < .01$. It can be concluded that intranational and cross-national studies show their own dynamics.

HYPOTHESIS TESTING

National differences in context variables (i.e., GNP, educational expenditure, latitude, and the four Hofstede dimension scores) were factor analyzed. Each publication constituted one case in the analysis, making a total of 37 (cross-national) comparisons. Two factors were extracted that explained 71% of the variance (eigenvalues of 3.25 and 1.69). The varimax-rotated loadings are represented in Table 2. The first factor referred to national differences in affluence; high loadings were found for GNP, educational expenditure (both

TABLE 1
Multiple Regression Analysis of Publication Year for Intranational
(*n* = 647) and Cross-National (*n* = 304) Comparisons

<i>Predictor</i>	<i>Standardized Regression Coefficient</i>	
	<i>Intranational Comparisons</i>	<i>Cross-National Comparisons</i>
Type of task ^a	0.28**	-0.01
Stimulus complexity ^b	0.04	-0.12*
Response complexity ^c	-0.12**	-0.08
Mean age	0.18**	0.08
Number of tasks administered	0.23**	0.25**
Sample size	0.05	0.36**
<i>R</i> ²	.15**	.22**

a. Coding of type of task: 1 = common Western task; 2 = locally developed Western task; 3 = locally developed non-Western task.

b. Stimulus complexity: 1 = simple reconstruction, substitution, speed and accuracy tasks; 2 = questions of factual knowledge, simple memory (e.g., digit span forward), drawing, simple arithmetic, free recall; 3 = embedded problems, block patterns, mazes, complex memory (digit span backward), "closure," jigsaw puzzles; 4 = sorting, classification, analysis, spatial orientation; 5 = insight, understanding, series extrapolations, analogies, and reasoning.

c. Response complexity: 1 = multiple choice, simple manual, simple writing; 2 = open end with simple answer, complex manual; 3 = free description.

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

per capita), latitude, individualism, and power distance. The second factor referred to national differences in masculinity and uncertainty avoidance. That these latter dimensions emerged as a single factor is a consequence of the relatively frequent comparisons of Japan on one hand, and the United States, Canada, and Western-European countries on the other hand. Japan has high scores on these dimensions compared to the other countries.

Factor scores of the two factors (for all independent cross-national comparisons) were then used as predictors of effect size in a multiple regression analysis. The multiple correlation was .38, $F(2, 258) = 22.26$, $p < .01$. The standardized regression coefficient of the affluence factor was highly significant, $\beta = 0.37$, $t(258) = 6.38$, $p < .01$. The regression coefficient of the second factor just failed to reach significance, $\beta = -0.11$, $t(258) = -1.959$, $p = .051$. Hypothesis 1, which stated that effect sizes are related to context variables, was supported (universalistic model).

The accuracy of the cumulative differences model was tested by correlating observed effect sizes to examinees' average age. For all studies combined, the correlation was positive and significant, as hypothesized, $r = .09$, $t(1,466) = 3.44$, $p < .01$. The correlation was .06, $t(1,137) = 1.87$, $p = .06$, for intranational studies and .20, $t(327) = 3.60$, $p < .01$, for cross-national studies. These

TABLE 2
Factor Loadings of Differences in Country Characteristics

National Differences	Factor	
	Affluence	Masculinity/ Uncertainty Avoidance
Capital expenditure on education	.77	.08
Latitude	.77	-.38
Gross National Product	.77	.03
Individualism	.85	.06
Power distance	.84	-.13
Masculinity	-.00	.87
Uncertainty avoidance	-.03	.90

latter two correlations differ significantly from each other, $z = 2.22, p < .01$. Although not documented here, correlations of about the same size were found when chronological age was replaced by educational age. In sum, the cumulative differences model could only be confirmed in cross-national studies.

According to the corollary of the cognitive complexity model, both stimulus and response complexity of tasks should be positively related to effect size. The correlation between stimulus complexity and effect size was positive and significant when intranational and cross-national studies were combined, $r = .10, t(1,553) = 3.91, p < .01$. The predicted relationship between effect size and response complexity was not observed, $r = .00, t(1,003) = .10, ns$. In intranational comparisons, the correlation for stimulus complexity was $.16, t(1,208) = 5.55, p < .01$, and $.07$ for response complexity, $t(691) = 1.80, ns$. Both correlations were nonsignificant in cross-national comparisons, $r = -.00, t(343) = -.01$, and $-.06, t(310) = -1.08$, respectively. So, the support for the cognitive complexity model was limited.

Evidence for the great divide model has frequently been derived from comparisons of literates and nonliterates. Unfortunately, the paucity of such comparisons in the present data did not warrant a direct hypothesis test. In addition, the high correlation between chronological and educational ages made it impossible to separate both. Therefore, we had to settle for a considerably weaker test. Only if the great divide model is taken also to apply to literate populations with a diverse educational background could the model be tested in the present data set. The hypothesis tested was whether tasks of abstract reasoning (operationalized here as the tasks with the highest complexity level, i.e., a stimulus complexity score of 5 as described before) have a higher effect size than tasks with lower complexity levels in cross-national

comparisons (i.e., the lowest four complexity levels). The prediction was not borne out. A t test showed that the averages of the more complex tasks ($M = 0.74$) and less complex tasks ($M = 0.70$) did not differ significantly, $t(343) = 0.45$, *ns*.

The influence of type of task on effect size was examined in an analysis of variance, with effect size as dependent variable and type of task as independent variable. The F ratio was significant, $F(2, 1,552) = 20.27$, $p < .01$. The mean values were .67 for common Western tasks, .58 for locally developed Western tasks, and .30 for locally developed non-Western tasks. Post hoc tests (Student-Newman-Keuls test with $\alpha = .05$) showed that all three means differed significantly from each other. The hypothesis stating that Western tasks yield larger cross-cultural differences than locally developed non-Western tasks was supported (bias model). Furthermore, the standard deviations of the effect sizes varied across the three types of tasks; the values were 0.61, 0.45, and 0.32 for common Western, locally developed Western, and locally developed non-Western tasks, respectively. In sum, effect sizes are smaller and more homogeneous for locally developed non-Western tasks than for common Western tasks.

Thus far, hypotheses were tested separately and a comparison of their relative importance has not yet been addressed. The latter question was addressed by carrying out multiple regression analyses for intranational and cross-national studies separately. Effect size was the dependent variable in both analyses (see Table 3). In the first analysis (involving the intranational comparisons), average age, type of task, stimulus complexity, and response complexity were independent variables. The analysis showed a modest though significant R^2 of .06, $F(3, 1,135) = 22.59$, $p < .01$; the type of task and the stimulus complexity had significant standardized regression coefficients, $\beta = -0.17$ and 0.15 , $t(1,135) = -6.05$ and 5.07 , respectively, $p < .01$ for both, whereas the coefficient of average age just failed to reach significance, $\beta = 0.06$, $t(1,135) = 1.93$, $p = .054$.

The second analysis considered cross-national comparisons. In addition to the predictors used in the first analysis, factor scores on the modernity and masculinity/uncertainty avoidance factors were included as independent variables. Two predictors were significant, namely, differences in affluence, $\beta = 0.32$, $t(245) = 5.49$, $p < .01$, and mean age, $\beta = 0.19$, $t(245) = 3.05$, $p < .01$; the regression coefficient of masculinity/uncertainty avoidance was just significant, $\beta = -0.12$, $t(245) = -1.98$, $p < .05$. Furthermore, the value of R^2 was .20, $F(5, 245) = 11.97$, $p < .01$. This value was significantly higher than the value of .06 found for intranational studies ($p < .01$). The major difference between the first and the second analysis is the type of predictor that was significant. Effect sizes were better predicted by task characteristics than by

TABLE 3
Multiple Regression Analysis of Effect Size for Intranational
(n = 1,139) and Cross-National Comparisons (n = 251)

Predictor	Standardized Regression Coefficient	
	Intranational Comparisons	Cross-National Comparisons
Type of task ^a	−0.17**	0.03
Stimulus complexity ^b	0.15**	−0.11
Mean age	0.06	0.19**
Differences in affluence	—	0.32**
Differences in masculinity/ uncertainty avoidance	—	−0.12
R ²	.06**	.20**

a. Coding of type of task: 1 = common Western task; 2 = locally developed Western task; 3 = locally developed non-Western task.

b. Stimulus complexity: 1 = simple reconstruction, substitution, speed and accuracy tasks; 2 = questions of factual knowledge, simple memory (e.g., digit span forward), drawing, simple arithmetic, free recall; 3 = embedded problems, block patterns, mazes, complex memory (digit span backward), “closure,” jigsaw puzzles; 4 = sorting, classification, analysis, spatial orientation; 5 = insight, understanding, series extrapolations, analogies, and reasoning.

**p < .01.

sample characteristics in cross-national comparisons, whereas the opposite was found in intranational studies.

The possibility was explored as to whether differences in regression coefficients had to be accounted for by the inclusion of additional predictors in the cross-national data set. It could be argued that multicollinearity of the predictors had suppressed the impact of task characteristics. Regression analyses were carried out separately for the intranational and cross-cultural data sets, in which effect sizes were predicted only by stimulus complexity and type of task. The regression equation of the intranational data set was $effect = 0.59 + 0.06 \times stimulus\ complexity - 0.13 \times type\ of\ task$, $R^2 = .05$; both regression coefficients, $t(1,207) = 5.30$ and -5.89 , and the multiple correlation, $F(2, 1,207) = 33.14$, are significant, all $ps < .01$. For the cross-national data set, the equation was $effect = 0.87 - 0.01 \times stimulus\ complexity - 0.09 \times type\ of\ task$, $R^2 = .004$; all statistics nonsignificant; regression coefficients, $t(342) = -0.35$ and -1.24 ; multiple correlation, $F(2, 342) = .77$; all $ps > .05$. A test of the equality of the regression lines revealed a highly significant result, $F(3, 1,549) = 10.35$, $p < .01$.

As another alternative interpretation, the shift from task to population characteristics in the regression analysis could be assumed to be induced by a restriction in range of the variables in either regression analysis. Inspection of the data did not support the plausibility of the reasoning; there were no

Figure 1: Distributions of Characteristics of Cross-National and Intranational Studies: (a) Number of Comparisons Per Study, (b) Sample Size, (c) Average Age

sizable differences in dispersion of any of the variables. For the dependent variable, the standard deviations in the intranational and cross-national regressions were, respectively, 0.50 and 0.58; for type of task, 0.62 and 0.52; for stimulus complexity, 1.26 and 1.41; and for age, 7.86 and 8.90. So, neither multicollinearity nor range restriction could account for the differences in intranational and cross-national regression equations. After having excluded the most obvious methodological explanations, a psychological interpretation of the shift becomes more plausible and will be explored in the next section.

Figure 2: Distributions of Country Characteristics: (a) Countries Involved in Cross-National Studies, (b) Countries in Intranational Studies, and (c) Weighted and Unweighted Gross National Product (GNP) Distribution in U.S. Dollars

DISCUSSION

The present study aimed at detecting patterns in cross-cultural studies on cognition reported between 1973 and 1994. The distribution of effect sizes showed a strong positive skewness. As a consequence, fairly distinct values

Figure 3: Distribution of Task Characteristics: (a) Most Frequently Applied Tests in Cross-National Studies; (b) Tests Applied in Intranational Studies, (c) Stimulus Complexity, and (d) Response Complexity

NOTE: CMMS = Columbia Mental Maturity Scale; K-ABC = Kaufman Assessment Battery for Children; McCarthy = McCarthy Scales of Children's Abilities; Raven = The Coloured and Standard Raven Test; SB = Stanford-Binet Intelligence Scale; WAIS(-R) = Wechsler Adult Intelligence Test and Wechsler Adult Intelligence Test Revised; WISC(-R) = Wechsler Intelligence Scale for Children and Wechsler Intelligence Scale for Children (Revised). Stimulus complexity: 1 = simple reconstruction, substitution, speed and accuracy tasks; 2 = questions of factual knowledge, simple memory (e.g., digit span forward), drawing, simple arithmetic, free recall; 3 = embedded problems, block patterns, mazes, complex memory (digit span backward), "closure," jigsaw puzzles; 4 = sorting, classification, analysis, spatial orientation; 5 = insight, understanding, series extrapolations, analogies, and reasoning. Response complexity: 1 = multiple choice, simple manual, simple writing; 2 = open end with simple answer, complex manual; 3 = free description.

Figure 4: Distribution of Effect Sizes of Cross-National and Intranational Studies

were obtained for central tendency measures such as the median and the mode. The latter was zero, implying that the absence of any cross-cultural difference is the most frequently observed finding in studies of mental tests. On the other hand, in Cohen's (1962) terminology, the median effect was medium-sized. Cross-national studies tended to report larger performance differences than did intranational studies. Five hypotheses derived from common models to explain cross-cultural differences in cognitive performance were tested. Cross-cultural performance differences were positively related to stimulus complexity of tasks, whereas differences in response complexity were unrelated to performance differences. This finding implies that the cognitive complexity model was supported only for stimulus complexity. Type of task was also related to observed performance differences as hypothesized by the bias model; common Western tasks showed the largest performance differences across groups, whereas locally developed non-Western tasks showed the smallest differences. National differences in afflu-

ence showed a positive relationship with performance differences of cultural groups, as predicted by the universalistic model. Furthermore, cross-cultural performance differences were positively related to years of schooling and age (which corroborated the cumulative differences model). Finally, the virtual absence of nonliterate subjects precluded a direct test of the great divide model. The weaker hypothesis that intergroup score differences in cross-national studies show larger performance differences in the area of abstract thinking than in other areas could not be confirmed.

Task characteristics were more powerful predictors of performance differences in intranational comparisons, whereas subject and group characteristics were more powerful in predicting cross-national performance differences. Statistical artifacts such as multicollinearity and restriction of range could not explain the differences. Therefore, a substantive reason is more likely to hold. We found a correlation of .37 between national differences in affluence (operationalized as the factor score on the first factor reported in Table 2) and performance, $t(259) = 6.34$, $p < .01$. Interestingly, a similar correlation was observed when studies of only the lowest level of stimulus complexity were considered, $r = .39$, $t(41) = 2.69$, $p < .01$. It appears that even cognitively simple tasks have characteristics that give rise to cross-national performance differences. It could be speculated that these include familiarity with stimuli, response procedures, and testing situations in general (Van de Vijver & Leung, 1997; Van de Vijver & Poortinga, 1997). In terms of the five models of cross-cultural performance differences, this interpretation is most in line with the bias model. Country characteristics such as affluence are successful predictors of cross-cultural performance differences, even for simple mental tasks, because they identify sources of bias in the instrument. It could be conjectured that an important country characteristic may be per capita expenditure on education that showed a high loading on the affluence factor. This interpretation is in line with findings that schooling has a rather general influence on the performance on cognitive tasks. Education prepares individuals to perform well on mental tasks by providing, among other things, exposure to and experience with educational tests and learning materials that are also found in mental tasks (Ceci, 1990; Rogoff, 1981; Rogoff & Chavajay, 1995). In sum, it is speculated here that national differences in affluence are related to differences in mental test performance because they are a proxy for educational differences and (somewhat fortuitously) identify sources of bias in the measures.

There were two problems in the application and interpretation of the results of the present meta-analysis. The first one is practical. A study could be included in the analysis only if the task and the subjects were described in sufficient detail and means and standard deviations were given. More than

three quarters of the cross-cultural studies of cognition that were examined did not meet this criterion, thereby possibly introducing an unknown sample bias in a meta-analysis. This complaint, often expressed in reports of meta-analyses, should be a concern to both authors and editors. Meta-analysis has become a popular tool during recent decades. Authors of empirical studies should anticipate inclusion of their work in secondary analyses.

The second problem is theoretical and more profound. The cross-cultural studies involving mental tasks that have been reported here can be seen as being derived from a hypothetical study in which all existing tasks have been administered to samples of all cultural groups. Unfortunately, the studies reported are a highly biased sample of the hypothetical study. For example, no comparison between literates and nonliterates could be made due to a lack of data in the present meta-analysis. There was an overrepresentation of subjects from affluent countries, both in cross-national and intranational studies (cf. Öngel & Smith, 1994). There was also a strong bias in tasks administered. The WISC and WISC-R were the most frequently administered instruments. Finally, the age distribution of the examinees showed an overrepresentation of children of primary school age. These biases have both theoretical and methodological implications. To start with the latter, figures such as correlations and effect sizes that are based on poorly represented cultures and tasks may be unstable, which may jeopardize the generalizability of the present findings. As for the theoretical implications, the influence of schooling on cognitive functioning may be somewhat overrated because of the overlap in school tasks and psychological tasks. Tasks with stimuli derived from daily life may well be less dependent on schooling as demonstrated in the daily cognition tradition (Sternberg & Wagner, 1986; Van de Vijver & Willemssen, 1993). The study of nonschooled or minimally schooled groups in both cross-national and intranational comparisons is needed to understand the role of cultural factors other than formal education in cognitive development.

By analyzing existing data, meta-analysis can help to define the scope of future studies. To deepen our insight into the relative importance of the various models proposed, experimental studies with more direct comparisons of models are required. For example, the present meta-analysis did not allow for a direct comparison of the relative importance of the cumulative deficit and bias models. It could well be that the two explanations are linked and that performance differences on cognitive tests increase with age because the tasks are closely linked to formal education, which shows an increasing difference with the age of the child. More powerful tests of models of intergroup differences in cognitive performance will require (quasi-)experimental designs that allow for more direct comparisons.

APPENDIX

STUDIES INCLUDED IN THE META-ANALYSIS

- Bagley, C. (1975). On the intellectual equality of races: Race and education across cultures. In G. K. Verma & C. Bagley (Eds.), *Race and education across cultures* (pp. 30-47). Heinemann: London.
- Bart, W. M., Kamal, A., & Lane, J. F. (1987). The development of proportional reasoning in Qatar. *Journal of Genetic Psychology*, 148, 95-103.
- Beauchamp, D. P., Samuels, D. D., & Griffiore, R. J. (1979). WISC-R information and Digit Span scores of American and Canadian children. *Applied Psychological Measurement*, 3, 231-236.
- Bentley, A. M. (1980). Pattern reproduction in two cultures. *International Journal of Psychology*, 15, 1-9.
- Bernal, E. M. (1984). Bias in mental testing. In C. R. Reynolds & R. T. Brown (Eds.), *Perspectives on bias in mental testing* (pp. 171-187). New York: Plenum.
- Bickersteth, P., & Das, J. P. (1981). Syllogistic reasoning among school children from Canada and Sierra Leone. *International Journal of Psychology*, 16, 1-11.
- Bleichrodt, N. (1989). *Intelligentiemeting ten behoeve van het onderwijs in Indonesië* [Intelligence measurement for education in Indonesia]. Amsterdam: Frisch.
- Board Schmeiser, C. (1982). Use of experimental design in statistical item bias studies. In R. A. Berk (Ed.), *Handbook of methods for detecting test bias* (pp. 64-95). Baltimore, MD: Johns Hopkins University Press.
- Boer, D. P., & Pugh, G. M. (1988). Canadian items for the WAIS-R Information subtest. *Canadian Journal of Behavioural Science*, 20, 273-286.
- Boivin, M. J. (1991). The effect of culture on a visual spatial memory task. *Journal of General Psychology*, 118, 327-334.
- Borkowski, J. G., & Krause, A. (1983). Racial differences in intelligence: The importance of the executive system. *Intelligence*, 7, 379-395.
- Born, M. P., & Lynn, R. (1994). Sex differences on the Dutch WISC-R: A comparison with the USA and Scotland. *Educational Psychology*, 14, 249-255.
- Bracken, B. A., Sabers, D. L., & Insko, W. (1987). Performance of Black and White children on the Bracken Basic Concept Scale. *Psychology in the Schools*, 24, 22-27.
- Bradley, R. H., Caldwell, B. M., Rock, J., Barnard, K. E., Gray, C., Hammond, M. A., Mitchell, S., Siegel, L., Rammey, C. T., Gottfried, A. W., & Johnson, D. L. (1988). Home environment and cognitive development in the first three years of life: A collaborative study involving six sites and three ethnic groups in North America. *Developmental Psychology*, 25, 217-235.
- Brandon, P. R., Newton, B. J., & Hammond, O. W. (1987). Children's mathematics achievement in Hawaii: Sex differences favoring girls. *American Educational Research Journal*, 24, 437-461.
- Brooks, I. R. (1976). Cognitive ability assessment with two New Zealand ethnic groups. *Journal of Cross-Cultural Psychology*, 7, 347-356.
- Brown, C. A., Keats, J. A., Keats D. M., & Seggie, I. (1980). Reasoning about implication: A comparison of Malaysian and Australian subjects. *Journal of Cross-Cultural Psychology*, 11, 395-410.
- Buium, N., & Turnure, J. E. (1977). A cross-cultural study of verbal elaboration productivity and memory in young children. *Child Development*, 48, 296-300.
- Burg, B., & Belmont, I. (1990). Mental abilities of children from different cultural backgrounds in Israel. *Journal of Cross-Cultural Psychology*, 21, 90-108.
- Calhoun, J. P. (1974). Developmental and sociocultural aspects of imagery in the picture word paired associate learning of children. *Developmental Psychology*, 10, 357-366.

- Carlson, D. B., LaBarba, R. C., Sclafani, J. D., & Bowers, C. A. (1986). Cognitive and motor development in infants of adolescent mothers: A longitudinal analysis. *International Journal of Behavioral Development*, 9, 1-13.
- Chandra, S. (1975). Cognitive development of Indians and Fijians. In J. W. Berry & W. J. Lonner (Eds.), *Applied cross-cultural psychology* (pp. 248-253). Lisse: Swets & Zeitlinger.
- Chissom, B. S., & Hoenes, R. L. (1976). A comparison of the ability of the D-48 Test and the IPAT Culture Fair Intelligence Test to predict SRA Achievement Test scores for 8th and 9th grade students. *Educational and Psychological Measurement*, 36, 561-564.
- Church, A. T., & Katigbak, M. S. (1987). Ecocultural bias in culture specific intelligence tests in an ecologically diverse culture: The Philippines. *Intelligence*, 11, 371-389.
- Claassen, N. (1990). The comparability of General Scholastic Aptitude Test across different population groups. *South African Journal of Psychology*, 20, 80-92.
- Clark, L. A., & Halford, G. S. (1983). Does cognitive style account for cultural differences in scholastic achievement? *Journal of Cross-Cultural Psychology*, 14, 279-296.
- Covin, T. M., Sattler, J. M., & Hilse, D. E. (1985). Comparison of California Test of Mental Maturity and WISC-R IQs among children suspected of being EMR or LD. *Psychology in the Schools*, 22, 43-45.
- Crawford, M. S., & Boer, D. P. (1985). Content bias in the WAIS-R Information subtest and some Canadian alternatives. *Canadian Journal of Behavioural Science*, 17, 79-86.
- Cudeck, R., & Claassen, N. (1983). Structural equivalence of an intelligence test for two language groups. *South African Journal of Psychology*, 13, 1-5.
- Dar, Y., & Resh, N. (1991). Socioeconomic and ethnic gaps in academic achievement in Israeli junior high schools. In N. Bleichrodt & P.J.D. Drenth (Eds.), *Contemporary issues in cross-cultural psychology* (pp. 322-333). Lisse: Swets & Zeitlinger.
- Davidson, G. R. (1974). Linguistic determinants of choice reaction time among Aborigines and White Australians. *Journal of Cross-Cultural Psychology*, 5, 199-211.
- Davies, I. R., & Corbett, G. G. (1991). Linguistic basicness and colour information processing. *International Journal of Psychology*, 26, 311-327.
- Davis, M., & Goodnow, J. J. (1977). Problem solving strategies used by Australian children with Australian and Greek parentage. *Journal of Cross-Cultural Psychology*, 8, 33-47.
- De Jong, M. J., & Van Batenburg, T. A. (1984). Etnische herkomst intelligentie en schoolkeuzeadvies [Ethnic origin, intelligence, and recommended school choice]. *Pedagogische Studiën*, 61, 362-371.
- De Lacey, P. R. (1972). A relationship between classificatory ability and verbal intelligence. *International Journal of Psychology*, 7, 243-246.
- Dean, R. S. (1980). Factor structure of the WISC-R with Anglos and Mexican Americans. *Journal of School Psychology*, 18, 234-239.
- Dillon, R., & Carlson, J. S. (1978). Testing for competence in three ethnic groups. *Educational and Psychological Measurement*, 38, 437-443.
- Douglas, J. D., & Wong, A. C. (1977). Formal operations: Age and sex differences in Chinese and American children. *Child Development*, 48, 689-692.
- DuBois, C. L., Sackett, P. R., Zedeck, S., & Fogli, L. (1993). Further exploration of typical and maximum performance criteria: Definitional issues, prediction, and White-Black differences. *Journal of Applied Psychology*, 78, 205-211.
- Engelhard, G. (1990). Gender differences in performance on mathematics items: Evidence from the United States and Thailand. *Contemporary Educational Psychology*, 15, 13-26.
- Entwisle, D. R., & Alexander, K. L. (1990). Beginning school math competence: Minority and majority comparisons. *Child Development*, 61, 454-471.

- Fahmeier, E. D. (1975). The effect of school attendance on intellectual development in Northern Nigeria. *Child Development*, 46, 281-285.
- Foorman, B. R., Yoshida, H., Swank, P. R., & Garson, J. (1989). Japanese and American children's styles of processing figural matrices. *Journal of Cross-Cultural Psychology*, 20, 263-295.
- Geary, D. C., & Whitworth, R. H. (1988). Dimensional structure of the WAIS-R: A simultaneous multi-sample analysis. *Educational and Psychological Measurement*, 48, 945-956.
- Gilleard, E., & Gilleard, C. (1989). A comparison of Turkish and Anglo American normative data on the Wechsler Memory Scale. *Journal of Clinical Psychology*, 45, 114-117.
- Ginsburg, H. P., Posner, J. K., & Russell, R. L. (1981). The development of knowledge concerning written arithmetic: A cross-cultural study. *International Journal of Psychology*, 16, 13-34.
- Ginsburg, H. P., Posner, J. K., & Russell, R. L. (1981). The development of mental addition as a function of schooling and culture. *Journal of Cross-Cultural Psychology*, 12, 163-178.
- Gonzales, R. R., & Roll, S. (1985). Relationship between acculturation cognitive style and intelligence: A cross-sectional study. *Journal of Cross-Cultural Psychology*, 16, 190-205.
- Grubb, H. J., & Ollendick, T. H. (1986). Cultural distance perspective: An exploratory analysis of its effect on learning and intelligence. *International Journal of Intercultural Relations*, 10, 399-414.
- Guthke, J., & Al Zoubi, A. (1987). Kulturspezifische Differenzen in den Coloured Progressive Matrices (CPM) und in einer Lerntestvariante der CPM [Culture-specific differences in the Coloured Progressive Matrices (CPM) and a learning potential version of the CPM]. *Psychologie in Erziehung und Unterricht*, 34, 306-311.
- Gutkin, T. B., & Reynolds, C. R. (1981). Factorial similarities of the WISC-R for White and Black children from the standardization sample. *Journal of Educational Psychology*, 73, 227-231.
- Halpin, G., Simpson, R. G., & Martin, S. L. (1990). An investigation of racial bias in the Peabody Picture Vocabulary Test Revised. *Educational and Psychological Measurement*, 50, 183-189.
- Hays, J. R., & Smith, A. L. (1980). Comparison of WISC-R and Culture-Fair Intelligence Test scores for three ethnic groups of juvenile delinquents. *Psychological Reports*, 46, 931-934.
- Hennessey, J. J., & Merrifield, P. R. (1978). Ethnicity and sex distinctions in patterns of aptitude factor scores in a sample of urban high school seniors. *American Educational Research Journal*, 15, 385-389.
- Hewstone, M., Hantzi, A., & Johnston, L. (1991). Social categorization and person memory: The pervasiveness of race as an organizing principle. *European Journal of Social Psychology*, 21, 517-528.
- Hickman, J. A., & Reynolds, C. R. (1986-1987). Are race differences in mental test scores an artifact of psychometric methods? A test of Harrington's experimental model. *Journal of Special Education*, 20, 409-430.
- Hilger, M. I., Klett, W. G., & Watson, C. G. (1976). Performance of Ainu and Japanese six year olds on the Goodenough Harris Drawing Test. *Perceptual and Motor Skills*, 42, 435-438.
- Holtzman, W. H. (Ed.). (1975). *Personality development in two cultures*. Austin: University of Texas Press.
- Irwin, M. H., Schafer, G. N., & Feiden, C. P. (1974). Emic and unfamiliar category sorting of Mano farmers and U.S. undergraduates. *Journal of Cross-Cultural Psychology*, 5, 407-423.
- Jachuck, K. (1982). Patterns of cognitive ability among boys in four sub-cultural groups. In R. Rath, H. S. Asthana, D. Sinha, & J.B.H. Sinha (Eds.), *Diversity and unity in cross-cultural psychology* (pp. 60-69). Lisse: Swets & Zeitlinger.
- Jahoda, G. (1979). On the nature of difficulties in spatial perceptual tasks: Ethnic and sex differences. *British Journal of Psychology*, 70, 351-363.

- Jensen, A. R. (1973). Level I and Level II abilities in three ethnic groups. *American Educational Research Journal*, 10, 263-276.
- Jensen, A. R. (1974). Interaction of Level I and Level II abilities with race and socioeconomic status. *Journal of Educational Psychology*, 66, 99-111.
- Jensen, A. R. (1977). An examination of culture bias in the Wonderlic Personnel Test. *Intelligence*, 1, 51-64.
- Jensen, A. R. (1984). The Black-White differences on the K-ABC: Implications for future tests. *Journal of Special Education*, 18, 377-408.
- Jensen, A. R. (1993). Spearman's hypothesis tested with chronometric information processing tasks. *Intelligence*, 17, 47-77.
- Jensen, A. R., & Figueroa, R. A. (1975). Forward and backward digit span interaction with race and I.Q.: Predictions from Jensen's theory. *Journal of Educational Psychology*, 8, 239-247.
- Jensen, A. R., & Inouye, A. R. (1980). Level I and Level II abilities in Asian White and Black children. *Intelligence*, 4, 41-89.
- Jensen, A. R., & McGurk, F. C. (1987). Black-White bias in "cultural" and "noncultural" test items. *Personality and Individual Differences*, 8, 295-301.
- Jensen, A. R., & Reynolds, C. R. (1982). Race, social class and ability patterns on the WISC-R. *Personality and Individual Differences*, 3, 423-438.
- Johnson, D. F., & Mihal, W. L. (1973). Performance of Blacks and Whites in computerized versus manual testing environments. *American Psychologist*, 28, 694-699.
- Johnson, D. L., Swank, P.-H., Virgil, M. B., Baldwin, C. D., Owen, M., & Luttman, D. (1993). Does HOME add to the prediction of child intelligence over and above SES? *Journal of Genetic Psychology*, 154, 33-40.
- Johnston, W. T., & Bolen, L. M. (1984). A comparison of the factor structure of the WISC-R for Blacks and Whites. *Psychology in the Schools*, 21, 42-44.
- Kalyan-Masih, V. (1989). Cognitive performance and cognitive styles of White American and non-White immigrant children. In D. M. Keats, D. Munro, & L. Mann (Eds.), *Heterogeneity in cross-cultural psychology* (pp. 261-273). Lisse: Swets & Zeitlinger.
- Kaniel, S., & Fisherman, S. (1991). Level of performance and distribution of errors in the Progressive Matrices test: A comparison of Ethiopian immigrant and native Israeli adolescents. *International Journal of Psychology*, 26, 25-33.
- Kaufman, A. S. (1984-1985). K-ABC and controversy. *Journal of Special Education*, 18, 409-444.
- Kaufman, A. S., & Doppelt, J. E. (1976). Analysis of WISC-R standardization data in terms of the stratification variables. *Child Development*, 47, 165-171.
- Kaufman, A. S., McLean, J. E., & Reynolds, C. R. (1988). Sex, race, residence, region and education differences on the 11 WAIS-R subtests. *Journal of Clinical Psychology*, 44, 231-248.
- Kearins, J. M. (1981). Visual spatial memory in Australian Aboriginal children of desert regions. *Cognitive Psychology*, 13, 434-460.
- Keats, D. M., & Fang, F.-X. (1992). The effect of modification of the cultural content of stimulus materials on social perspective taking ability in Chinese and Australian children. In S. Iwawaki, Y. Kashima, & K. Leung (Eds.), *Innovations in cross-cultural psychology* (pp. 319-327). Lisse: Swets & Zeitlinger.
- Kee, D. W., Siaw, S. N., & Carter, L. (1985). Population differences in constructive memory. Children's long term memory for noun pairs. *Contemporary Educational Psychology*, 10, 268-274.
- Klich, L. Z., & Davidson, G. R. (1983). A cultural difference in visual memory: On le voit on ne le voit plus. *International Journal of Psychology*, 18, 189-201.

- Klippel, M. D. (1975). Measurement of intelligence among three New Zealand ethnic groups. *Journal of Cross-Cultural Psychology*, 6, 365-376.
- Knapp, P. A., & Seagrim, G. N. (1981). Visual memory in Australian Aboriginal children and children of European descent. *International Journal of Psychology*, 16, 213-231.
- Kok, F. (1988). *Vraagpartijdigheid* [Item bias]. Amsterdam: De Amstel.
- Krohn, E. J., & Lamp, R. E. (1989). Concurrent validity of the Stanford Binet Fourth Edition and K-ABC for Head Start children. *Journal of School Psychology*, 27, 59-67.
- Kunen, S., Cohen, R., & Solman, R. (1981). A levels of processing analysis of Bloom's taxonomy. *Journal of Educational Psychology*, 73, 202-211.
- Laughon, P., & Torgesen, J. K. (1985). Effects of alternate testing procedures on two subtests of the Woodcock Johnson Psycho-Educational Battery. *Psychology in the Schools*, 22, 160-163.
- Li, C., & Shallcross, D. J. (1992). The effect of the assumed boundary in the solving of the nine dot problem on a sample of Chinese and American students 6-18 years old. *Journal of Creative Behavior*, 26, 53-64.
- Lieblich, A., Kugelmass, S., & Erlich, C. (1975). Patterns of intellectual ability in Jewish and Arab children in Israel II: Urban matched samples. *Journal of Cross-Cultural Psychology*, 6, 218-226.
- Little, A. (1975). The educational achievement of ethnic minority children in London schools. Race and education across cultures. In G. K. Verma & C. Bagley (Eds.), *Race and education across cultures* (pp. 48-69). London: Heinemann.
- Lynn, R., Chan, J.W.C., & Eysenck, H. J. (1991). Reaction times and intelligence in Chinese and British children. *Perceptual and Motor Skills*, 72, 443-452.
- Lynn, R., & Hampson, S. (1986). Intellectual abilities of Japanese children: An assessment of 2 1/2 - 8 1/2 year olds derived from the McCarthy Scales of Children's Abilities. *Intelligence*, 10, 41-58.
- Lynn, R., & Hampson, S. (1987). Further evidence on the cognitive abilities of the Japanese: Data from the WPPSI. *International Journal of Behavioral Development*, 10, 23-36.
- Lynn, R., & Pagliari, C., & Chan, J. (1988). Intelligence in Hong Kong measured for Spearman's *g* and the visuospatial and verbal primaries. *Intelligence*, 12, 423-433.
- Lynn, R., & Shigehisa, T. (1991). Reaction times and intelligence: A comparison of Japanese and British children. *Journal of Biosocial Science*, 23, 409-416.
- Marjoribanks, K. (1978). Ethnicity school attitudes intelligence and academic achievement. *International Journal of Psychology*, 13, 167-178.
- Marjoribanks, K. (1979). Ethclass, the achievement syndrome and children's cognitive performance. *Journal of Educational Research*, 72, 327-333.
- Marjoribanks, K. (1991). Relationship of children's ethnicity, gender, and social status to their family environments and school related outcomes. *Journal of Social Psychology*, 131, 83-91.
- Marmorale, A. M., & Brown, F. (1982). Constancy of WISC IQs of Puerto Rican White and Black children. In L. L. Adler (Ed.), *Cross-cultural research at issue* (pp. 137-143). New York: Academic Press.
- Marwit, S. J., & Neumann, G. (1974). Black and White children's comprehension of standard and nonstandard English passages. *Journal of Educational Psychology*, 66, 329-332.
- Matarazzo, J. D., & Wiens, A. N. (1977). Black intelligence test of cultural homogeneity and Wechsler Adult Intelligence Scale scores of Black and White police applicants. *Journal of Applied Psychology*, 62, 57-63.
- Matthtis, F. E., Spooner, W. E., Coble, C. R., & Takemura, S. (1992). A study of the logical thinking skills and integrated process skills of junior high school students in North Carolina and Japan. *Science Education*, 76, 211-222.
- McCauley, D. E., & Colberg, M. (1983). Transportability of deductive measurement across cultures. *Journal of Educational Measurement*, 20, 81-92.

- Meacham, J. A. (1975). Patterns of memory abilities in two cultures. *Developmental Psychology*, 11, 50-53.
- Mehryar, A. H., Tashakkori, A. Y., & Farideh, K. F. (1987). The application of the Goodenough Harris Draw A Man test to a group of Iranian children in the city of Shiraz. *British Journal of Educational Psychology*, 57, 401-406.
- Miele, F. (1979). Cultural bias in the WISC. *Intelligence*, 3, 149-164.
- Misawa, G., Motegi, M., Fujita, K., & Hattori, K. (1984). A comparative study of intellectual abilities of Japanese and American children on the Columbia Mental Maturity Scale (CMMS). *Personality and Individual Differences*, 5, 173-181.
- Miura, I. T., Okamoto, Y. K., Chungsoon, C., Steere, M., & Fayol, M. (1993). First graders' cognitive representation of number and understanding of place value: Cross-national comparisons: France, Japan, Korea, Sweden, and the United States. *Journal of Educational Psychology*, 85, 24-30.
- Montie, J. E., & Fagan, J. F. (1988). Racial differences in IQ: Item analysis of the Stanford Binet at 3 years. *Intelligence*, 12, 315-332.
- Moore, K. A., & Snyder, N. O. (1991). Cognitive attainment among firstborn children of adolescent mothers. *American Sociological Review*, 56, 612-624.
- Mundy Castle, A., Wilson, D. J., Sibanda, P. S., & Sibanda, J. S. (1989). Cognitive effects of LOGO among Black and White Zimbabwean girls and boys. *International Journal of Psychology*, 24, 539-546.
- Muse, C. (1989). A study of Samoan children on 11 cognitive tasks. In D. M. Topping, D. C. Crowell, & V. N. Kobayashi (Eds.), *Thinking across cultures: The third international conference on thinking* (pp. 43-54). Hillsdale, NJ: Lawrence Erlbaum.
- Naglieri, J. A., & Jensen, A. R. (1987). Comparison of Black-White differences on the WISC-R and the K-ABC: Spearman's hypothesis. *Intelligence*, 11, 21-43.
- Nagoshi, C. T., & Johnson, R. C. (1987). Cognitive abilities profiles of Caucasian vs. Japanese subjects in the Hawaii Family Study of Cognition. *Personality and Individual Differences*, 8, 581-583.
- Nagoshi, C. T., Johnson, R. C., & Honbo, K. A. (1992). Assortative mating for cognitive abilities, personality, and attitudes: Offspring from the Hawaii Family Study of Cognition. *Personality and Individual Differences*, 13, 883-891.
- Nenty, H., & Johnson, D.T.E. (1981). A cross-cultural analysis of the fairness of the Cattell Culture Fair Intelligence Test using the Rasch model. *Applied Psychological Measurement*, 5, 355-368.
- Newcombe, N., Rogoff, B., & Kagan, J. (1977). Developmental changes in recognition memory for pictures of objects and scenes. *Developmental Psychology*, 13, 337-341.
- Orasanu, J., Lee, & Scribner, S. (1979). The development of category organization and free recall: Ethnic and economic group comparisons. *Child Development*, 50, 1100-1109.
- Owen, K. (1992). The suitability of Raven's Standard Progressive Matrices for various groups in South Africa. *Personality and Individual Differences*, 13, 149-159.
- Parker, J. F. (1977). Free recall of abstract and concrete words by American and Ghanaian college students. *International Journal of Psychology*, 12, 243-252.
- Parmar, R. S. (1989). Cross-cultural transfer of nonverbal intelligence tests: An (in)validation study. *British Journal of Educational Psychology*, 59, 378-388.
- Petitto, A. L., & Ginsburg, H. P. (1982). Mental arithmetic in Africa and America: Strategies principles and explanations. *International Journal of Psychology*, 17, 81-102.
- Poortinga, Y. H., & Van der Flier, H. (1988). The meaning of item bias in ability tests. In S. H. Irvine & J. W. Berry (Eds.), *Human abilities in cultural context* (pp. 166-183). Cambridge, UK: Cambridge University Press.

- Powers, S., & Barkan, J. H. (1986). Concurrent validity of the Standard Progressive Matrices for Hispanic and non-Hispanic seventh grade students. *Psychology in the Schools*, 23, 333-336.
- Price Williams, D. R., & Ramirez M. (1977). Divergent thinking, cultural differences, and bilingualism. *Journal of Social Psychology*, 103, 3-11.
- Pugh, G. M., & Boer, D. P. (1991). Normative data on the validity of Canadian substitute items for the WAIS-R Information subtest. *Canadian Journal of Behavioural Science*, 23, 149-158.
- Reschly, D. J., & Sabers, D. L. (1979). Analysis of test bias in four groups with the regression definition. *Journal of Educational Measurement*, 16, 1-9.
- Reynolds, C. R. (1980). Differential construct validity of intelligence as popularly measured: Correlations of age with raw scores on the WISC-R for Blacks Whites males and females. *Intelligence*, 4, 371-379.
- Reynolds, C. R. (1980). An examination for test bias in a preschool battery across race and sex. *Journal of Educational Measurement*, 17, 137-146.
- Reynolds, C. R., & Gutkin, T.B.A. (1981). A multivariate comparison of the intellectual performance of Black and White children matched on four demographic variables. *Personality and Individual Differences*, 2, 175-180.
- Reynolds, C. R., & Jensen, A. R. (1983). WISC-R subscale patterns of abilities of Blacks and Whites matched on full scale IQ. *Journal of Educational Psychology*, 75, 207-214.
- Rogoff, B., & Waddell, K. J. (1982). Memory for information organized in a scene by children from two cultures. *Child Development*, 53, 1224-1228.
- Rushton, J. P. (1992). Life history comparisons between Orientals and Whites at a Canadian university. *Personality and Individual Differences*, 13, 439-442.
- Saccuzzo, D. P., Johnson, N. E., & Russell, G. (1992). Verbal versus performance IQs for gifted African American, Caucasian, Filipino, and Hispanic children. *Psychological Assessment*, 4, 239-244.
- Samuel, W. (1977). Observed IQ as a function of test atmosphere, tester expectation, and race of tester: A replication for female subjects. *Journal of Educational Psychology*, 69, 593-604.
- Samuel, W. (1983). Sex differences in spatial ability reflected in performance on IQ subtests by Black or White examinees. *Personality and Individual Differences*, 4, 219-221.
- Samuel, W. L., Soto, D., Parks, M., Ngissah, P., & Jones, B. (1976). Motivation, race, social class and IQ. *Journal of Educational Psychology*, 68, 273-285.
- Sandoval, J. (1979). The WISC-R and internal evidence of test bias with minority groups. *Journal of Consulting and Clinical Psychology*, 47, 919-927.
- Sandoval, J. (1982). The WISC-R factorial validity from minority groups and Spearman's hypothesis. *Journal of School Psychology*, 20, 198-204.
- Sattler, J. M. (1982). *Assessment of children's intelligence and special abilities*. Boston: Allyn & Bacon.
- Schooler, D. L., & Anderson, R. L. (1979). Race differences on the Developmental Test of Visual Motor Integration, the Slosson Intelligence Test, and the ABC Inventory. *Psychology in the Schools*, 16, 453-456.
- Schratz, M. M. (1978). A developmental investigation of sex differences in spatial (visual analytic) and mathematical skills in three ethnic groups. *Developmental Psychology*, 14, 263-267.
- Scott, R. (1976). Race of examiner as a variable in test attainments of preschool children. *Perceptual and Motor Skills*, 42, 1167-1173.
- Shepherd, J. W., Deregowski, J. B., & Ellis, H. D. (1974). A cross-cultural study of recognition memory for faces. *International Journal of Psychology*, 9, 205-212.

- Shigehisa, T., & Lynn, R. (1991). Reaction times and intelligence in Japanese children. *International Journal of Psychology*, 26, 195-202.
- Sijtsma, K., & Resing, W.C.M. (1991). Scalability of an intelligence test for different ethnic groups. In N. Bleichrodt & P.J.D. Drenth (Eds.), *Contemporary issues in cross-cultural psychology* (pp. 361-373). Lisse, the Netherlands: Swets & Zeitlinger.
- Silverstein, A. B. (1973). Factor structure of the Wechsler Intelligence Scale for Children for three ethnic groups. *Journal of Educational Psychology*, 65, 408-410.
- Smirni, P., Villardita, C., & Zappala, G. (1983). Spatial memory span in adolescents: Cultural and sex differences. *Perceptual and Motor Skills*, 57, 175-178.
- Solkoff, N. (1974). Race of examiner and performance on the Wechsler Intelligence Scale for Children: A replication. *Perceptual and Motor Skills*, 39, 1063-1066.
- Speidel, G. E., Forran, D. C., & Jorden, C. (1989). On the learning and thinking styles of Hawaiian children. In D. M. Topping, D. C. Crowell, & V. N. Kobayashi (Eds.), *Thinking across cultures: The third international conference on thinking* (pp. 55-78). Hillsdale, NJ: Lawrence Erlbaum.
- Stevenson, H. W., Stigler, J. W., Lee, S., Lucker, G. W., Kitamura, S., & Hsu, C. (1985). Cognitive performance and academic achievement of Japanese, Chinese and American children. *Child Development*, 56, 718-734.
- Stigler, J. W., Lee, S. Y., & Stevenson, H. W. (1986). Digit memory in Chinese and English: Evidence for a temporally limited store. *Cognition*, 23, 1-20.
- Stone, B. J. (1992). Prediction of achievement by Asian American and White children. *Journal of School Psychology*, 30, 91-99.
- Sung, Y. H., Dawis, R., V. (1981). Level and factor structure differences in selected abilities across race and sex groups. *Journal of Applied Psychology*, 66, 613-624.
- Svanum, S., & Bringle, R. G. (1982). Race social class and predictive bias: An evaluation using the WISC, WRAT and teacher ratings. *Intelligence*, 6, 275-286.
- Takeuchi, M., & Scott, R. (1992). Cognitive profiles of Japanese and Canadian kindergarten and first grade children. *Journal of Social Psychology*, 132, 505-512.
- Taylor, L. J., & Skanes, G. (1976). Cognitive abilities in Inuit and White children from similar environments. *Canadian Journal of Behavioural Science*, 8, 1-7.
- Taylor, L. J., & Skanes, G. R. (1976). Level I and Level II intelligence in Inuit and White children from similar environments. *Journal of Cross-Cultural Psychology*, 7, 157-168.
- Taylor, L. J., & Skanes, G. R. (1977). A cross-cultural examination of some of Jensen's hypotheses. *Canadian Journal of Behavioural Science*, 9, 315-322.
- Taylor, R. L., & Partenio, I. (1984). Ethnic differences on the Bender Gestalt: Relative effects of measured intelligence. *Journal of Consulting and Clinical Psychology*, 52, 784-788.
- Taylor, R. L., & Richards, S. B. (1991). Patterns of intellectual differences of Black, Hispanic, and White children. *Psychology in the Schools*, 28, 5-9.
- Taylor, R. L., Ziegler, E. W., & Partenio, I. (1984). An investigation of WISC-R verbal performance differences as a function of ethnic status. *Psychology in the Schools*, 21, 437-441.
- Turnage, T. W., & McGinnies, E. (1973). A cross-cultural comparison of the effects of presentation mode and meaningfulness on short term recall. *American Journal of Psychology*, 86, 369-381.
- Umansky, W., & Cohen, L. R. (1980). Race and sex differences on the McCarthy Screening Test. *Psychology in the Schools*, 17, 400-404.
- Valencia, R. R. (1979). Comparison of intellectual performance of Chicano and Anglo third grade boys on the Raven Coloured Progressive Matrices. *Psychology in the Schools*, 16, 448-453.

- Valencia, R. R., & Rankin, R. J. (1985). Evidence of content bias on the McCarthy Scales with Mexican American children: Implications for test translation and nonbiased assessment. *Journal of Educational Psychology*, 77, 197-207.
- Valencia, R. R., & Rankin, R. J. (1986). Factor analysis of the K-ABC for groups of Anglo and Mexican American children. *Journal of Educational Measurement*, 23, 209-219.
- Van de Langeberg, G. (1989). *Prestatieverbeteringen op reactietijd taken van uiteenlopende cognitieve complexiteit: Een cross-cultureel onderzoek* [Performance increments on reaction time tasks of different levels of cognitive complexity]. Unpublished master's thesis, Tilburg University.
- Van de Rijt, B. (1990). *Reactiesnelheidstest: Een aanvulling voor allochtonen op de bestaande intelligentietests* [Reaction time test: A complement for migrants to existing intelligence tests]. Unpublished master's thesis, Tilburg University.
- Van de Vijver, F.J.R. (1991). *Inductive thinking across cultures: An empirical study*. Helmond: WIBRO.
- Van de Vijver, F. J., Daal, M., & Van Zonneveld, R. (1986). The trainability of formal thinking: A cross-cultural comparison. *International Journal of Psychology*, 21, 589-615.
- Van der Flier, H. (1980). *Vergelijkbaarheid van individuele testprestaties* [Comparability of individual test performances]. Lisse, the Netherlands: Swets & Zeitlinger.
- Vance, H. B., Hankins, N., & McGee, H. (1979). A preliminary study of Black and White differences on the revised Wechsler Intelligence Scale for Children. *Journal of Clinical Psychology*, 35, 815-819.
- Vane, J. R. (1973). Intelligence and achievement test results of kindergarten age children in England, Ireland and the United States. *Journal of Clinical Psychology*, 29, 191-193.
- Vernon, P. E., & Jensen, A. R. (1984). Individual differences in intelligence and speed of information processing. *Personality and Individual Differences*, 5, 411-423.
- Verster, J. M. (1973). A study of intellectual structure in two groups of South African scientists. *Psychologia Africana*, 15, 169-190.
- Verster, J. M., & Prinsloo, R. J. (1988). The diminishing test performance gap between English speakers and Afrikaans speakers in South Africa. In S. H. Irvine & J. W. Berry (Eds.), *Human abilities in cultural context* (pp. 534-559). Cambridge, UK: Cambridge University Press.
- Victoria, M. D., Victoria, C. G., & Barros, F. C. (1990). Cross-cultural differences in developmental rates: A comparison between British and Brazilian children. *Child: Care, Health and Development*, 16, 151-164.
- Vincent, K. R., & Cox, J. A. (1974). A re-evaluation of Raven's Standard Progressive Matrices. *Journal of Psychology*, 88, 299-303.
- Vries, A. K., & De Bunjes, L.A.C. (1989). Zwakke rekenprestaties bij buitenlandse adoptie-kinderen in de basisschool [Poor mathematics results of foreign adopted children in primary education]. *Kind en Adolescent*, 10, 165-174.
- Wagner, D. A. (1974). The development of short term and incidental memory: A cross-cultural study. *Child Development*, 45, 389-396.
- Warden, P. G., & Prawat, R. S. (1975). Convergent and divergent thinking in Black and White children of high and low socioeconomic status. *Psychological Reports*, 36, 715-718.
- Weinberg, R. A., Scarr, S., & Waldman, I. (1992). The Minnesota Transracial Adoption Study: A follow up of intelligence test performance at adolescence. *Intelligence*, 16, 117-135.
- West, A. M., Mackintosh, N. J., & Mascie, T.C.G. (1992). Cognitive and educational attainment in different ethnic groups. *Journal of Biosocial Science*, 24, 539-554.
- Whitworth, R. H., & Chrisman, S. M. (1987). Validation of the Kaufman Assessment Battery for Children comparing Anglo and Mexican American preschoolers. *Educational and Psychological Measurement*, 47, 695-702.

- Willemse, G. (1989). *Keuzereactietijdtaken in multiculturele context* [Choice reaction time measures in multicultural context]. Unpublished master's thesis, Tilburg University.
- Wilson, D., Mundy Castle, A., & Sibanda, P. (1991). Cognitive effects of LOGO and computer-aided instruction among Black and White Zimbabwean primary school girls. *Journal of Social Psychology, 131*, 107-116.
- Wing, H. (1980). Profiles of cognitive ability of different racial/ethnic and sex groups on a multiple abilities test battery. *Journal of Applied Psychology, 65*, 289-298.
- Yule, W., Berger, M., Rutter, M., & Yule, B. (1975). Children of West Indian immigrants II. Intellectual performance and reading attainment. *Journal of Child Psychology and Psychiatry, 16*, 1-17.
- Zeidner, M. (1987). Gender and cultural interaction effects on scholastic aptitude test performance: Some Israeli findings. *International Journal of Psychology, 22*, 111-119.
- Zeidner, M. (1987). Sociocultural group differences in the level and pattern of scholastic aptitude test scores: The Israeli scene. *International Journal of Psychology, 22*, 363-373.

NOTE

1. The year 1986 was somewhat arbitrarily chosen as a year close to the middle of the period of coverage for which data of many countries were available.

REFERENCES

- Ballantine, J. H. (1993). *The sociology of education: A systematic analysis* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Berk, R. A. (Ed.). (1982). *Handbook of methods for detecting test bias*. Baltimore, MD: Johns Hopkins University Press.
- Berry, J. W., Poortinga, Y. H., Segall, M. H., & Dasen, P. R., (1992). *Cross-cultural psychology: Research and applications*. Cambridge, UK: Cambridge University Press.
- Born, M. P., Bleichrodt, N., & Van der Flier, H. (1987). Cross-cultural comparison of sex-related differences on intelligence tests: A meta-analysis. *Journal of Cross-Cultural Psychology, 18*, 283-314.
- Buck-Morss, S. (1975). Socio-economic bias in Piaget's theory and its implications for cross-cultural studies. *Human Development, 18*, 35-49.
- Camilli, G., & Shepard, L. N. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, UK: Cambridge University Press.
- Ceci, S. J. (1990). *On intelligence . . . more or less: A bio-ecological treatise on intellectual development*. Englewood Cliffs, NJ: Prentice Hall.
- Chapman, M. (1988). Contextuality and directionality of cognitive development. *Human Development, 31*, 92-106.
- Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology, 65*, 145-153.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

- Hofstede, G. (1980). *Culture's consequences*. Beverly Hills, CA: Sage.
- Holland, P. W., & Wainer, H. (Eds.). (1993). *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum.
- Irvine, S. H. (1979). The place of factor analysis in cross-cultural methodology and its contribution to cognitive theory. In L. Eckensberger, W. Lonner, & Y. H. Poortinga (Eds.), *Cross-cultural contributions to psychology* (pp. 300-341). Lisse, the Netherlands: Swets & Zeitlinger.
- Irvine, S. H., & Berry, J. W. (1988). The abilities of mankind: A revaluation. In S. H. Irvine & J. W. Berry (Eds.), *Human abilities in cultural context* (pp. 3-59). Cambridge, UK: Cambridge University Press.
- Jensen, A. R. (1977). Cumulative deficit in intelligence of Blacks in the rural South. *Developmental Psychology*, 13, 184-191.
- Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.
- Mercer, J. R. (1984). What is a racially and culturally nondiscriminatory test? In C. R. Reynolds & R. T. Brown (Eds.), *Perspectives on bias in mental testing* (pp. 293-356). New York: Plenum.
- Öngel, Ü., & Smith, P. B. (1994). Who are we and where are we going? *JCCP* approaches its 100th issue. *Journal of Cross-Cultural Psychology*, 25, 25-53.
- Rogoff, B. (1981). Schooling and the development of cognitive skills. In H. C. Triandis & A. Heron (Eds.), *Handbook of cross-cultural psychology* (Vol. 4, pp. 233-294). Boston: Allyn & Bacon.
- Rogoff, B., & Chavajay, P. (1995). What's become of research on the cultural basis of cognitive development? *American Psychologist*, 50, 859-877.
- Segall, M. H., Dasen, P. R., Berry, J. W., & Poortinga, Y. H. (1990). *Human behavior in global perspective*. New York: Pergamon.
- Spearman, C. (1923). *The abilities of man*. New York: Macmillan.
- Sternberg, R. J., & Wagner, R. K. (Eds.). (1986). *Practical intelligence*. Cambridge, UK: Cambridge University Press.
- Tulviste, P. (1991). *The cultural-historical development of verbal thinking*. Commack, NY: Nova Science.
- United Nations. (1988). *1985/1986 statistical yearbook*. New York: Author.
- Van de Vijver, F.J.R., & Leung, K. (1997). Methods and data analysis of comparative research. In J. W. Berry, Y. H. Poortinga, & J. Pandey (Eds.), *Handbook of cross-cultural psychology* (2nd ed., vol. 1, pp. 257-300). Chicago: Allyn & Bacon.
- Van de Vijver, F.J.R., & Poortinga, Y. H. (1997). Towards an integrated analysis of bias in cross-cultural assessment. *European Journal of Psychological Assessment*, 13, 29-37.
- Van de Vijver, F.J.R., & Willemssen, M. E. (1993). Abstract thinking. In J. Altarriba (Ed.), *Culture and cognition* (pp. 317-342). Amsterdam: North Holland.
- Van der Flier, H. (1980). *Vergelijkbaarheid van individuele testprestaties* [Comparability of individual test performances]. Lisse, the Netherlands: Swets & Zeitlinger.
- Vernon, P. E. (1950). *The structure of human abilities*. New York: John Wiley.
- Vernon, P. E. (1979). *Intelligence: Heredity and environment*. San Francisco: Freeman.

Fons van de Vijver teaches cross-cultural psychology and statistics at Tilburg University, the Netherlands. His main interests are cross-cultural studies of cognition and methodological aspects of cross-cultural comparisons. Together with Kwok Leung (Chinese University of Hong Kong), he has written a chapter in the recent Handbook of Cross-Cultural Psychology and a book on research methodology in cross-cultural psychology (Methods and Data Analysis for Cross-Cultural Research, 1997, Sage).