European Journal of Personality
Eur. J. Pers. **21**: 667–706 (2007)
Published online in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/per.634

# The *g*-Factor of International Cognitive Ability Comparisons: The Homogeneity of Results in PISA, TIMSS, PIRLS and IQ-Tests Across Nations

HEINER RINDERMANN*
*Institute of Psychology, Otto-von-Guericke-University Magdeburg, Germany*

*Abstract*

*International cognitive ability and achievement comparisons stem from different research traditions. But analyses at the interindividual data level show that they share a common positive manifold. Correlations of national ability means are even higher to very high (within student assessment studies,* r = *.60–.98; between different student assessment studies [PISA-sum with TIMSS-sum]* r = *.82–.83; student assessment sum with intelligence tests,* r = *.85–.86). Results of factor analyses indicate a strong* g*-factor of differences between nations (variance explained by the first unrotated factor: 94–95%). Causes of the high correlations are seen in the similarities of tests within studies, in the similarities of the cognitive demands for tasks from different tests, and in the common developmental factors at the individual and national levels including known environmental and unknown genetic influences. Copyright © 2007 John Wiley & Sons, Ltd.*

Key words: intelligence; knowledge; education; student assessment; educational achievement; OECD; IEA

## INTRODUCTION

The past century of empirical research on cognitive abilities has convincingly demonstrated their relevance for both individuals and societies across a wide variety of criteria. For individuals, both IQ (Schmidt & Hunter, 2004) and literacy (OECD, 2000) have been linked with job performance, longevity (Gottfredson & Deary, 2004; OECD, 2000), and low criminality (Herrnstein & Murray, 1994). Moreover, performance on Piagetian tasks is correlated with moral development (Piaget, 1932). At the level of nations, economic wealth has been linked with both IQ (Lynn & Vanhanen, 2006) and with Third International Mathematics and Science Study (1994) and Trends in International

*Correspondence to: Dr habil. Heiner Rindermann, Institute of Psychology, Otto-von-Guericke-University Magdeburg, PO Box 4120, D-39016 Magdeburg, Germany.
E-mail: heiner.rindermann@gse-w.uni-magdeburg.de

*Received 29 November 2006*
*Revised 13 February 2007*
*Accepted 13 February 2007*

Mathematics and Science Study (1999 and 2003; TIMSS for both) results (Hanushek & Kimko, 2000). The positive impact of cognitive abilities—the correlations are usually interpreted as mainly caused by effects of cognitive ability—is seen in different paradigms of cognitive ability research, in psychometric intelligence research, in students' academic performance research and in cognitive development research. Further, the impact is seen at the levels of both individuals and societies.

In addition, results from different intelligence and student performance tests—as indicators for successful cognitive development—appear to depend on highly similar conditions. What furthers intelligence and cognitive development also helps to develop student academic performance. For example, attendance of classes and schools with high ability classmates is positively related to students' cognitive performance and increase on literacy or similar tests (Ammermüller & Pischke, 2006; Bishop, 2004; Fertig, 2002; Hanushek, Kain, Markman & Rivkin, 2003; Tiedemann & Billmann-Mahecha, 2004) and IQ (Rindermann, 2007); and the education of both parents and upbringing in an intact family predict and seem to support cognitive performance (in schools, Dornbusch, Ritter, Leiderman, Roberts & Fraleigh, 1987; for TIMSS, Gonzales, 2000; for PISA 2000, Gaeth, 2005, Fertig, 2002; for PISA 2003, Prenzel et al., 2004; for IQ, Armor, 2003, Meisenberg, Lawless, Lambert & Newton, 2006).

The empirical evidence for the homogeneity of cognitive abilities (*g*-factor, Spearman, 1904; Jensen, 1998) is thus very strong. But the conceptual and developmental associations between intelligence and student achievement have been neglected up to now in international student assessment studies. Especially at the level of international and cross-cultural comparisons, the relationships between different measures of cognitive ability (including achievement) have been largely disregarded. This is because these international and cross-cultural comparisons of cognitive competences stem from different research traditions, are domiciled in different disciplines, use different measurement methods (e.g. paper–pencil tests with short or long items, observation of behaviour in everyday situations or in experiments) and different statistical analysis methods (e.g. classical test theory or item-response-theory, factor analysis or model tests) and have developed different scales with different content (verbal, math, science, problem solving, space, etc.) and with different proximity to general or school-specific knowledge. All these differences make it scientifically and academically difficult to notice the similarities.

The thesis of this paper is that student achievement assessments and intelligence tests primarily measure a common cognitive ability *at the macro-social level*. This ability consists of the ability to think (intelligence) and of knowledge (extent of true and relevant knowledge, the ability to acquire and use knowledge). Intelligence and knowledge are measured to different degrees by student assessment and intelligence tests; the intelligence-knowledge-differences are greater within different intelligence tests and within student assessment tests than between student assessment and intelligence tests. In detail, the research questions addressed here are as follows:

1. Are the cross-national correlations between different scales of student cognitive ability high enough to justify the assumption of a strong *g*-factor?
2. At the macro-social level, is it more appropriate to work with a single scale, a single study, *or* with sum values of different scales and studies?
3. Do content analyses justify the assumption of homogenous cognitive demands across different scales within and between different cognitive ability and achievement tests?



DOI: 10.1002/per

4. What are possible causes of a *g*-factor at the macro-social level?

I will first describe the most relevant cognitive ability studies, and then discuss numerous methodological problems involved in answering the above research questions.

## SOURCES OF INTERNATIONAL COGNITIVE ABILITY COMPARISONS

### International student assessment studies

Among international cognitive ability comparisons, student assessment studies have attracted the most attention in recent decades. Since the International Association for the Evaluation of Educational Achievement (IEA) Reading-Study of 1991, many countries have participated in Reading-, Math- and Science-Studies (International Assessment of Educational Progress [IAEP]-II 1991, TIMSS 1995, TIMSS 1999, TIMSS 2003, Programme for International Student Assessment [PISA] 2000, PISA 2003, Progress in International Reading Literacy Study [PIRLS] 2001). In comparison to studies using intelligence tests and Piagetian measures, these student assessment studies have the most representative and largest samples of well-defined populations using consistent instruments and tasks across nations. But the samples only cover persons actually attending schools—pupils. In countries with high non-attendance rates at different ages (from 9/10 to 15 and 17/18/19 years) like Mexico, Brazil or Albania, or even South Korea and Austria, the results overestimate the average ability for the whole population of youth. The studies are briefly summarised below.

In the *IEA-Reading-Study* 1991 (Elley, 1992), grade levels with mostly 9- and 14-year-old pupils in 26 and 30 nations were targeted for testing. The response rates of pupils ranged from 59% (Nigeria) to 100% (Botswana, Finland, Hong Kong, Iceland, Singapore, Slovenia, Spain and Hungary). The study was age oriented, but age requirements were not met in at least some nations: 9-year-old pupils were 10.7 years old in Venezuela, and 14-year-old pupils were 15.5 years old in Zimbabwe. The study was designed to measure *reading literacy*, 'the ability to understand and use those written language forms required by society and/or valued by the individual' (Elley, 1992, p. 3). The test material consisted of narrative and expository documents and charts, tables, maps, graphs and lists, drawn from typical home, school, society or work contexts.

In the two *IAEP-II Mathematics and Science* Studies 1991 (Lapointe, Askew & Mead, 1992; Lapointe, Mead & Askew, 1992), mostly 9- and 13-year-old pupils in 13 and 19 nations, respectively, were tested. The samples in many countries (China, Italy, Israel, Mozambique) were not representative (e.g. only 4% of the pupils in Italy of the defined age group), the participation rates were low, and the country sample was small. Therefore, this study is not used here for aggregation of a general cognitive ability factor across nations.

The *TIMS-Studies* (1994–95, 1999, 2003) measure cognitive competences and knowledge in mathematics and sciences ('mathematics and science literacy achievement'; Mullis, Martin, Beaton, Gonzalez, Kelly, & Smith, 1998, p. 3) in a total of 63 nations. The studies were organised by the IEA and were (mainly) grade-level oriented. In 1994–95 grades 4, 8 and 12 were tested (in some countries grades 3 and 7). For grade 12 the attendance rates were poor and varied widely among countries (e.g. Greece 10%, Latvia 3%; Mullis et al., 1998, p. 18; see Bracey, 2000); therefore, they are not analysed here. In the other studies participation rates vary between 45 and 100%. Age differences of up to 2 years between pupils in different countries (e.g. TIMSS 1998, 8th grade: Greece, 13.6 years; Columbia, 15.7 years) constitute the main problem of all grade-level studies.



DOI: 10.1002/per

TIMSS measured math and science in shorter items (than PISA) and was oriented by school curricula. Many tasks consisted of short knowledge questions (e.g. why does a candle go out under a cover, fossil fuels arise from what?).

In the *PISA-Studies* (2000–2002 and 2003), 15-year-old pupils were measured in reading literacy, mathematical literacy, science literacy and, in 2003, in problem solving in a total of 48 nations. The studies were organised by the OECD and were age-level oriented. Persons not attending school were not tested. The participation rates for schools and pupils were good in most countries, but there were large differences in school attendance rates (e.g. in Albania only 43% of the 15-year-old youth were pupils; in Mexico, 52%; in Brazil, 53%; and in Peru, 66%; OECD, 2001, p. 232 and OECD, 2003a, p. 251). The tasks were not oriented by national school curricula. Instead, the goal was to measure reading, mathematics, science and problem-solving *literacy*:

– *Reading literacy*: 'Reading literacy is understanding, using and reflecting on written texts, in order to achieve one's goals, to develop one's knowledge and to participate in society' (OECD, 2003b, p. 108).
– *Mathematical literacy*: 'Mathematical literacy is an individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgements and to use and engage with mathematics in ways that meet the needs of the individual's life as a constructive, concerned and reflective citizen' (OECD, 2003b, p. 24).
– *Scientific literacy*: 'Scientific literacy is the capacity to use scientific knowledge, to identify questions and to draw evidence-based conclusions in order to understand and help make decisions about the natural world and the changes made to it through human activity' (OECD, 2003b, p. 133).
– *Problem solving*: 'Problem solving is an individual's capacity to use cognitive processes to confront and resolve real, cross-disciplinary situations where the solution path is not immediately obvious and where the literacy domains or curricular areas that might be applicable are not within a single domain of mathematics, science or reading' (OECD, 2003b, p. 156).

The tasks were usually long; pupils had to read texts, tables and graphs and could find the answers in the majority of the tasks without using additional knowledge.

In *PIRLS* (Mullis, Martin, Gonzales, & Kennedy, 2003), pupils of the 4th grade were tested in *reading literacy,* or 'the ability to understand and use those written language forms required by society and/or valued by the individual. Young readers can construct meaning from a variety of texts (p. 33)'. This kind of task is similar to reading (and some mathematics, science and problem solving) tasks in PISA and IEA-Reading studies, but age-adapted to be easier (i.e. with shorter texts, use of common and less abstract words, less cognitive complex demand). In the 2001 study 33 nations took part. The participation rates of pupils in schools were always higher than 84%.

In comparison to usual items in intelligence tests, the items in the students' achievement tests were *longer*, with *more text*, and in TIMSS there were some pure knowledge questions. Pure and short figural tasks like in the Raven's Standard Progressive Matrices (SPM) were not included. The majority of IEA-Reading, PISA and PIRLS tasks can be solved with the information given in the tasks. Intelligence researchers such as Gottfredson (1997) judge literacy tasks (e.g. National Adult Literacy Survey, NALS) as intelligence tasks; the same position is held by educational researchers such as Prais (2003, p. 143): 'Answering such questions [PISA, race-task] correctly may be more a test of 'common



DOI: 10.1002/per

sense', or of 'IQ', than the results of mathematical schooling at this age'. Scholastic Assessment Test (SAT or 'SAT Reasoning Test') results are also commonly used as indicators of intelligence (Kanazawa, 2006a).

The results from the student assessment studies are usually standardised ($M = 500$ and $SD = 100$); the reference population for PISA is the collection of OECD-states.

## International intelligence test studies

Only one international comparison study has been carried out using a uniform intelligence test measured over a short time period under more or less standardised conditions. This is the study with the Cattell Culture Fair Test 3 (CFT3) non-verbal scale (Buj, 1981), probably conducted in the 1970s in 21 European countries and Ghana. The tests were administered in capital cities or in the biggest town in each country. But researchers believe the data from this study are of dubious quality: nobody knows the author; he did not work at a university; the way he collected so much data is unknown; the description of samples and testing procedure is scanty; and only one single two-page-long publication exists. The correlations with other measures, except PISA, are good (see below).

The test data collection of Lynn and Vanhanen (2002, 2006; see Figure 2b) is now well known. They collected data from 113 countries (in 2002, 81 countries) and estimated missing test information for an additional 79 countries (in 2002, 104 countries). The correlation between the 2002 estimated and the 2006 measured values is $r = .92$ ($N = 28$). Differently from the student assessment studies and from Buj, Lynn and Vanhanen relied on published and unpublished results from (more or less) representative samples that took a range of common intelligence tests, incorporating them all into one standard scale (mean for Great Britain was 100, standard deviation of Great Britain was 15, reference year 1979).

The tests typically used were the SPM and Colored Progressive Matrices (CPM), less often the Advanced Progessive Matrices (APM), Wechsler Intelligence Scale for Children (WISC), CFT, and, more rarely, the American Otis Test, Kaufman-Assessment Battery for Children (K-ABC), Draw-a-Man-Test, McCarthy Test and others. For some countries the results of Buj (1981) were used, too. When possible, results were taken from different studies and averaged. Correlations of different test scores (within countries) across countries were very high ($r = .92$–$.95$; Lynn & Vanhanen, 2006, p. 62).

In comparison to the student assessment studies, the representativeness and comparability of the samples used in these studies of intelligence test results is low (e.g. Hunt & Wittmann, in press). Test data from 79 countries are still missing, the collection of test results is an ongoing process. The used tests included not only tests without any school content like SPM, CPM, APM and CFT, but also tests like the WISC that contain verbal and number tasks and knowledge questions. Also the measurement years vary, which poses problems due to secular trends toward higher achievement in IQ-tests (Flynn, 1987). The standard adjustment of two (Raven Matrices) or three (all other tests) IQ-points per decade is probably not always correct, because since the 1970s the secular rise of intelligence in developing countries has been higher than in First World countries (Meisenberg, Lawless, Lambert, & Newton, 2005, Meisenberg et al., 2006). Missing values for countries with unknown test results were estimated using the means for neighbouring countries. This is not necessarily correct as there is some evidence that the absence of test results is correlated with social factors impeding cognitive development (e.g. war or poverty). In addition some errors in data have been observed (Loehlin, 2007; Mackintosh, 2007). The mixture of tests and the not always clear representativeness of



DOI: 10.1002/per

samples seem to be the most serious problems. Nevertheless, the IQ-test data collection has advantages because it includes not only pupils in school but also older people, and it includes more developing countries.

### Cross-cultural studies in other traditions (Piaget, problem solving, behaviour in everyday situations)

Jean Piaget's theory of cognitive development stimulated a lot of cross-cultural research (e.g. Greenfield, 1966; Berry & Dasen, 1974; Dasen, 1977; Hallpike, 1978; Oesterdiekhoff, 2000). The used tasks in this research are not paper–pencil tasks, but tasks with real objects, and people are observed in daily behaviour. Their solutions and comments during problem solving are used as data. This research has produced quantitative data (e.g. percentage of young people or people at a given age located on the concrete-operational thinking or formal-operational thinking level) which correspond to results from psychometric tests (e.g. at macro-social level Lebanon and USA: Za'Rour, 1971), but data sets for a large nation sample have either not been collected or are unknown.

Similar problems exist with tasks of complex problem solving (Badke-Schaub & Strohschneider, 1998; Strohschneider & Güss, 1998, 1999). Results for university students from India, Brazil and Germany correspond to national intelligence and student assessment test results, but the samples within the countries and the country samples for international comparisons were too small and were not representative.

Quantitative data is missing also for the observation of cognitive behaviour in everyday situations (Dagona, 1994; Gordon, 1997; Hallpike, 1978; Lévy-Bruhl, 1966; Oesterdiekhoff, 1992). Qualitative descriptions, analyses and explanations of rational and irrational behaviour in different cultures and historic eras such as the Middle Ages help to reveal differences in cognitive abilities across cultures and to recognise their relevance for everyday behaviour. Moreover, observations of cognitive behaviour in everyday situations are a further source for underpinning the validity of cognitive differences between cultures: just as the ultimate indicator of football playing competence is not shooting goals in table football or target shot, but playing good football in a real-life play within a team against an opponent team, the ultimate indicator of intelligence is cognitive operation in everyday situations. Although observations of everyday behaviour (or indirect by the use of documents and artifacts) could be transformed into quantitative data (e.g. historiometric analyses by Simonton, 2006), no such data seem presently available for countries.

## METHODOLOGICAL PROBLEMS

International student assessment and intelligence test studies suffer from different problems of sample representativeness and statistical methods; mainly the comparability across nations is questioned (e.g. Collani, 2001; Prais, 2003; Wuttke, 2006). The criticisms are not new.[1] The main methodological problems and their solution in the present analysis are described below.

[1]For example 30 years ago the important mathematics educationist Freudenthal (1975) wrote that, 'Backed by international authority sensational comparisons were published, which in fact were nothing but artefacts of invalid instruments' (p. 176). About the quality of the organisation of these studies, he wrote: 'The communication between Headquarters and periphery was bad. Decisions made at Headquarters did not reach the periphery, or were misunderstood, reactions at the periphery were misunderstood at Headquarters,' (p. 178); and very sarcastically: 'IEA, id est absurdum' (p. 164).



DOI: 10.1002/per

## Nation

The definition of the term 'nation' (or 'country') is sometimes unclear. In some studies information is presented for only parts of nations, or only some regions of the nations took part (e.g. England, Wales, Scotland or Northern Ireland vs. United Kingdom; British Columbia vs. Canada; German, French or Italian regions of Switzerland vs. Switzerland; French or Flemish parts of Belgium vs. Belgium). Some countries or regions are not UN members or are not independent (Hong Kong, Taiwan, Palestine); others are very small (Liechtenstein). I used the data in the following way: if values are presented only for regions of a country but not for the whole, I aggregated them if major regions took part (e.g. German *and* French parts of Switzerland for Switzerland; England *and* Scotland of the United Kingdom for the UK; French *and* Flemish parts of Belgium for Belgium); if only England of the UK participated, this was taken for UK, but British Columbia as a small part of Canada was not taken for Canada (thus, Canada was excluded). All mentioned territories and small countries (Hong Kong, Taiwan, Liechtenstein) used in student assessment studies or in Lynn and Vanhanen's (2002, 2006) intelligence test studies were included except for Palestine, which took part only in TIMSS 2003.

## Participation rates and conditions

In many countries the school attendance and participation rates are problematic (i.e. are not representative, provide fragmentary information, or contain questionable data):

1. Not all young persons attend school.
2. Not all pupils attend regularly school.
3. The attendance of school will depend on educational, regional and economic circumstances of the pupil and his or her parents including violence, war and problems of transportation.
4. The studies are easier to implement in well-organised regions and schools of a country.
5. The national study organisers can exclude pupils from participation (i.e. mentally retarded pupils, pupils at special schools, pupils in lower grades in PISA). The exclusion rates differ among the countries (Prais, 2003; Neuwirth, Ponocny, & Grossmann, 2006; Wuttke, 2006). For example, Prais (2003, p. 149) mentions critically about UK and Germany, 'How it came about that two OECD countries could interpret instructions from OECD HQ [headquarters] in such opposite ways may be left as an exercise for the student of Kafkaism'.
6. The male–female ratio of participating pupils differs among the countries (PISA 2003: in South Korea 41% were female, in France, 53%; Wuttke, 2006, p. 111; Neuwirth et al., 2006).
7. Response patterns or the test administration and handling of response patterns differ among the countries (e.g. guessing, making two crosses in multiple choice tasks when only one is expected; Wuttke, 2006, p. 113, 135).
8. Schools and pupils rejected participation to different degrees (depending on country, school form and integration into the educational system). Because better and longer schooled youth and youth with increased cognitive development due to better familial and environmental conditions were more likely to participte, results in many countries were overestimated. This will lead to underestimation of the correlations between the



DOI: 10.1002/per

dimensions within one study or between the studies at the individual and the macro-social data level.

In some studies participation rates were documented, but could have different meanings:

1. Percentage of invited pupils who participated.
2. Percentage of invited schools that participated.
3. Percentage of youth attending school.

Low participation rates will result in lower representativeness of results and over-estimations of results for pupils and all youth. The information given in the reports about participation rates is not always correct; for example, sometimes rates higher then 100% are documented (Wuttke, 2006, p. 106).

An additional problem is the adherence to age requirements. In the IEA-Reading and PISA studies, pupils of a well-defined age period were targeted. But in the Reading-Study 1991 there were large differences among countries in the ages of pupils actually tested; in PISA the age interval ranged strictly from 15 years and 3 months to 16 years and 2 months. In TIMSS and PIRLS, the sample definitions consisted of grade and age requirement mixtures (grades and complete classes with pupils of typical ages 10, 14 etc.). But grade is not a consistently defined term across countries. For example, in the Netherlands 'school' starts with age 4 and in Finland with 7, but in the Netherlands pupils do perhaps the same things in school before age 6 as children in Finland's kindergarten: drawing and some training with numbers, words and letters. Strictly age-oriented international comparisons are based on clearer definitions of the samples, making the comparisons easier and more convincing. Even more importantly, ability is best defined as achievement at a given age. This means that ability is age-dependent (a given performance at a younger age means higher ability), and quality of education is demonstrated by the capability of its students to demonstrate high ability at young ages.

This all makes it necessary to adjust the results from individual countries if they are to be used as representative indicators of the abilities of pupils in schools or for the whole (young or even adult) populations of countries. But the problem remains that information about the conditions of participation and about participation rates are sometimes missing or not very reliable (Barro & Lee, 1993, even report fabrication of educational data in some countries). Adjustments are only possible when information is available about the nature of the adjustments necessary. The specific adjustments themselves are also questionable. Because of this, I report analyses both with raw data and adjusted data, and compare them. The easiest way to reduce study-specific errors is the use of aggregated values from different studies.

## Adjustments for improving representativeness and comparability

Adjustments are not easy because any adjustment presupposes some kind of knowledge or assumption about the direction and the extent of necessary adjustment. For example, if children are older in one country than in another, a downward adjustment is necessary for that country because of biological maturation, because of learning inside and outside school, and because ability is defined as performance at a given age. If the school attendance rate of the youth is lower, the correction has to be downwards (if taking the result as an indicator of the entire youth population), because school attendance is the



DOI: 10.1002/per

single most important factor for cognitive development (Ceci, 1991; Greenfield, 1966; Lurija, 1976). Pupils and schools not participating in a student assessment study probably have fewer abilities because the factors leading to absence or test rejection also tend to lead to lower cognitive ability development (e.g. irregular school attendance, low compliance with educational requirements in school and family). Appropriate adjustments differ depending on whether the ability results should be representative only for pupils or for all youth of a society. The quality of an educational system is defined in part by its capability to reach the entire youth. The society consists of all, not only of selected youth. Therefore, the aim is to have results representative of all youth. Adjustments are never perfect; they are based on plausibility when some information is available. I used the following corrections for scale and sum values:

*IEA-Reading-Study 1991 (Elley, 1992)*

Adjustment formulas for pupils older or younger than the targeted age and for low participation rates of pupils were as follows: (formula for the 9-year-old pupil study: $\text{Read09k} = \text{Read09} - ((\text{Age9–9.7692}) \times 42) - ((100\text{-PR9E}) \times 2)$; for the 14-year-old pupil study: $\text{Read14k} = \text{Read14} - ((\text{Age14–14.7333}) \times 42) - ((100\text{-PR14E}) \times 2)$). Read09 is the country ability mean in the '9'-study, and Read14 is the country ability mean in the '14'-study. The country mean age in the '9'-study is 9.7692, the country mean age in the '14'-study is 14.7333. The results of applying the formulas were that: countries with pupils 1 year older than the mean had 42 points ($d = .42$) subtracted, countries with pupils 1 year younger than the mean had 42 points added. One year of school thus represented a gain of about 42 points.

This choice was based on the following considerations. Intelligence test studies report an improvement of about IQ = 3 points for 1 year of school attendance (on a scale with a mean of 500, this is 20 points; Winship & Korenman, 1997), but student assessment studies have reported higher gains: in mathematics from grades 3 to 4, about 60 points (Mullis, Martin, Beaton, Gonzalez, Kelly, & Smith, 1997, p. 31); in mathematics from grades 4 to 8, about 30 points each year (Mullis et al., 1997, p. 43); in science from grades 3 to 4, about 55 points (Martin, Mullis, Beaton, Gonzalez, Smith, & Kelly, 1997, p. 29); in science from grades 4 to 8, about 40 points each year (Martin et al., 1997, p. 41), in mathematics from grades 7 to 8 about 30 points (Beaton, Mullis, Martin, Gonzalez, Kelly, & Smith,1996, p. 29); in science from grades 7 to 8, about 35 points (Beaton, Martin, Mullis, Gonzalez, Smith, & Kelly, 1996, p. 29). The overall mean of these gains is 42 points. It is possible that school attendance prompts more ability increase in lower grades and at low ability levels.

Although an adjustment of 42 points ($d = .42$) is higher than the results indicated in intelligence research ($d = .20$; Winship & Korenman, 1997; Ceci, 1991), there are some reasons for higher corrections. The 42-point difference stems from different countries with large samples; in tests using with curricula-related topics like TIMSS, higher growth can be expected; and greater than targeted mean ages are associated with many other negative educational characteristics like low school attendance rate. In studies using tests less closely related to school content and in studies of older pupils like PISA, the correction should be lower (e.g. 35 points per year for Germany, DPK, 2005, p. 5).

Countries with a 90% pupil participation rate are adjusted downward by 20 points, while those with an 80% pupil participation rate are adjusted downward by 40 points. If half of a generation does not attend school, the result will be adjusted downward by $d = 1$. Plausibility analyses will be presented below.



DOI: 10.1002/per

*TIMSS 1995 (Beaton, Mullis, 1996; Martin et al., 1997; Mullis et al., 1997)*
The formula for adjusting for older or younger than mean pupil age (formula for the 8th grade was: TI895k = TI895−((AgeTi958–14.308) × 42); for 4th grade it was TI495k = TI495−((AgeTi954–10.232) × 42)). The results of applying the adjustment formulas were that countries with pupils 1 year older than the mean (14.308 or 10.232 years) were adjusted downward by 42 points, while countries with pupils 1 year younger than the mean were adjusted upward by 42 points. For the 8th grade, there was also information about participation rates ('coverage of 13-year-old students'; Beaton, Mullis, et al., 1996, p. A12) and the correction formula was: TI895kk = TI895k−((100-ParRT958) × 2). The results of applying the adjustment were that countries with a 90% pupil participation rate had 20 points subtracted, and those with a 80% pupil participation rate 40 had points subtracted. For Kuwait and Israel no information about attendance rates was provided, so the value of the next Muslim neighbour Iran (72%) was used for Kuwait, the worldwide mean (87.57%) was used for Israel.

*TIMSS 1999 (Martin et al., 2000; Mullis et al., 2000)*
The formula for adjusting for older or younger than mean pupil age was TI899k = TI899−((AgeTi998–14.366) × 42). The results after applying the adjustment were that countries with pupils 1 year older than the mean (14.366 years) were adjusted downward by 42 points, and countries with pupils 1 year younger than the mean were adjusted upward by 42 points. No information was presented in reports about participation rates.

*TIMSS 2003 (Martin, Mullis, Gonzalez, & Chrostowski, 2004; Mullis, Martin, Gonzalez, & Chrostowski, 2004)*
The formula for adjusting for older or younger than mean pupil age was TI803k = TI803−((AgeTi038–14.459) × 42) for the 8th grade and TI403−((AgeTi034–10.368) × 42) for the 4th grade. The results of applying the adjustment formula were that countries with pupils 1 year older than the mean (14.459 or 10.368 years) had 42 points subtracted, and countries with pupils 1 year younger than mean had 42 points added. No information was presented in the reports about participation rates.

*PISA 2000 (OECD, 2003a)*
The samples from 2000 and later ($N = 41$) were used. Country sub-samples (e.g. Belgium and Switzerland) were not used. Because of the strict age requirements in this study, no age adjustments were made. However, low school attendance rates were adjusted. The formula used was PISA00k = PISA00−((100-PRP00) × 2). The results of the adjustment were that countries with 90% attendance rate had 20 points subtracted; those with 80% attendance rate had 40 points subtracted. Other adjustments are conceivable.

The adjustment formula used is based on the assumption that the scores follow a normal distribution from which the lower part has been deleted. For example, when the lowest 16% of a normally distributed sample with a mean IQ of 100 are missing, the mean resulting IQ will be 104–105 (or 529 on the scale used in PISA). The formula used subtracts 4.8 IQ-points or 32 points on the PISA scale from the measured average and leads to marginally larger adjustments than a formula that is based on the assumption that the presented scores follow a normal distribution. This is particularly true for extremely low school enrolment ratios, e.g. for 50% missing, the formula subtracts 15 IQ-points or 100 PISA-points, although in a normally distributed sample the resulting mean would be only about 12 IQ-points or 80 PISA-points higher. The assumption is that with very low

 
DOI: 10.1002/per

school enrolment ratios, further factors such as poor nutrition and health care, low appreciation of education and of reading books, argumentation and thinking in everyday life, and intellectual simplicity of home environments and society will act to reduce IQ.

*PISA 2003 (OECD, 2004a, 2004b)*
Because the age requirements in this study were strict, no age adjustments were made. Low school attendance rates were corrected by the formula PISA03K = PISA03 − ((100-PartRP03) × 2). The results of this formula were that countries with 90% attendance rate had 20 points subtracted, with 80% attendance rate, 40 points were subtracted.

*PIRLS 2001 (Mullis et al., 2003)*
The formula for adjusting for older or younger than mean pupil age was PIRLSK = PIRLS − ((Age–10.312) × 42). The results of this adjustment were that countries with pupils 1 year older than the mean (10.312 years) had 42 points subtracted, and countries with pupils 1 year younger than the mean had 42 points added. No information was presented in the reports about participation rates.

*IQ (Lynn & Vanhanen, 2006)*
The actual IQ-data included 113 countries, and there were estimated data for an additional 79 countries (see Figure 2b). The correlation between the 2002 estimated and 2006 measured data was $r = .92$ ($N = 28$). Lynn and Vanhanen estimated IQs for 79 countries by using the mean of neighbouring countries with a similar population (race). But the absence of IQ measurement information is not accidental. It is more likely in countries with no or only a few universities and research institutions, with less developed educational systems, with less cultural interest in cognitive development, and with difficult social research conditions like war, poverty and no liberty (e.g. Afghanistan, Angola, Bangladesh, Haiti), all circumstances that do not indicate positive conditions for cognitive development. In small countries (e.g. Latvia, Luxembourg, Cyprus), IQ-data were missing probably only because of the small populations. Because of this, the adjustment was conditioned on lack of participation in student assessment studies. For the estimated data, five IQ-points were subtracted. Some smaller countries that did not participate in IQ and student assessment studies may still have been underestimated (e.g. Andorra, Bahamas). For Liechtenstein, which is missed completely in Lynn and Vanhanen (not measured, not estimated, but participated in PISA), the mean of Switzerland, Austria and Germany was used.

All the following analyses were done with both unadjusted and adjusted ability data.

## Aggregation within and across studies

The aggregation of different ability tests within one study (in each year and grade or age interval) is possible by using the arithmetic mean without standardising scores (e.g. for PISA 2003 and its four scales). But aggregation of scales across different studies is possible only after standardisation because different scales have different means and standard deviations (e.g. 500 and 100 vs. 100 and 15), and the same tests have different standardisation samples for different ages or grade levels and in different years (e.g. IEA-Reading, TIMSS and PISA, or TIMSS 1995 grade 4 vs. 8; PISA 2000 vs. 2003).

The following procedure was applied: First, the different tests of one study and age or grade level were aggregated (e.g. for TIMSS 1995 8th grade: mathematics and science). Second, after combined calculation of the means and standard deviations by using the



DOI: 10.1002/per

participating countries in two studies, I aggregated different age or grade levels of each study in each year (e.g. IEA-Reading 1991, 9- and 14-years old; TIMSS 1995 4th and 8th grade). Third, after combined calculation of the means and standard deviations by using the participating countries in (at least) two studies, I aggregated the different survey years of each study type (PISA 2000 and 2003, or TIMSS 1995, 1999 and 2003). Fourth, after renewed standardisation (see above), I separately aggregated all age-oriented studies (PISA) and all grade-oriented studies (TIMSS and PIRLS) because of their different approach toward age. Fifth, after renewed standardisation, I obtained a total average for all student assessment studies (age-oriented studies, grade-oriented studies, IEA-Reading 1991). The IEA-Reading study was not included within (fourth) the age-oriented studies because age requirements were not met. This study was included only with the (fifth) *all student assessment studies*. Here, age-oriented studies (PISA) and grade-oriented studies (TIMSS and PIRLS) were given a double weight because the IEA-Reading 1991 study is older, had only one measurement point, employed only one scale, and many study requirements were not met.

Finally (sixth), I calculated one total score for *all cognitive ability studies* (IQ, student assessment studies). Student assessment studies were given a double weight in this total because they have newer and larger samples and they consist of more cross-national studies.

Aggregation has two important advantages: (1) Aggregation helps to reduce study- and scale-specific biases and unreliabilities. (2) Aggregation increases the country data sample size (especially for developing countries); consequently, the analysis includes samples that are more representative for more regions of the world. The disadvantage is that country values are based on differently large samples of persons, from different years and studies. At the macro-social level investigating subjects such as the secular increase in intelligence or developmental effects between wealth and cognitive ability, the sum value of the IQ-collection (*all cognitive ability studies*) is not appropriate, because IQ-data stem from different decades. Depending on the time period researched, the aggregated value across all the student assessment studies would be better, as it reflects ability level over a narrower time span, the transition to the 21st century ('millennial cognitive ability').

All aggregation was done twice, for unadjusted and for adjusted data (see section on data adjustment above).

## Further cognitive ability indicators and attributes of societies

For comparisons were also used indicators of cognitive abilities from older, smaller or less representative studies: (1) The mean of the *OECD-Adult-Literacy-Study* in 20 nations (Murray, Kirsch & Jenkins, 1997). (2) Intelligence-test results presented by *Buj* (1981, $N = 22$). (3) *'Intellectual' order of the immigrants* in the USA 1921 (Shiraev and Levy, 2004, p. 131, source National Academy of Sciences, $N = 14$ mainly European countries). (4) The mean average percent correct of the *IAEP-II-studies* 1991 in mathematics and science for 9- and 13-year-old pupils (sum $N = 19$ countries; Lapointe, Askew, & Mead, 1992; Lapointe, Mean, & Askew, 1992; NCES, 1992). (5) The mean of old student assessment studies from 1964 to 1972 in 19 countries collected by Lee and Barro (1997).

For the *educational level of adults* three variables were aggregated: (1) Rate of literate adults in 1991 from Vanhanen (1997) in 172 nations. (2) Rate of persons between 12 and 19 years old from 1960 to 1985 (in the 1990s adults) having graduated secondary school, $N = 117$ (Mankiw, Romer, & Weil, 1992). (3) Years of school attendance in 1990, 1995 and



DOI: 10.1002/per

2000 of persons 25 years old or older from Barro and Lee (2000; $N = 107$). This sum value of education exists for 173 countries (mean computed after standardisation, $\alpha = .94$).

*Wealth* was measured by the gross national product (GNP; purchasing power parity) per capita 1998 from Lynn and Vanhanen (2002), $N = 185$; their sources were UNO data sets.

Indicators for *democracy* 1950–2004 stem from Vanhanen (2005) and Marshall and Jaggers (2000), $\alpha = .95$, $N = 183$.

## RESULTS

The data were analyzed at the national level (differences between countries and cultures). Only manifest correlations are reported. As data were used unadjusted and adjusted national values.

### Correlations among different scales of the same student assessment studies in the same year

In the *TIMS Studies*, correlations with unadjusted data between mathematics and science were high (see Table 1), ranging from .87–.97. In the *PISA Studies*, correlations between reading, mathematics, science and problem solving were also high, ranging from .95–.99. The correlations with unadjusted data within the studies between different scales were extremely high (mean $r = .95$). In the adjusted data they were even higher (mean $r = .97$; see Table 1 bottom left). Correlations between unadjusted and adjusted scales were always higher than $r = .90$. The lowest was $r = .91$ in the IEA-Reading study from 1991.

This means that in countries where pupils are good in mathematics, they are also good in sciences, and where they are good in sciences, they are good in reading and in problem solving etc.

### Correlations among scales and studies of same student assessment in different grades, ages, or years

TIMSS 1994, 1999 and 2003 and PISA 2000 and 2003 enable a comparison of correlation patterns (see Table 1). The correlations in *TIMSS* within mathematics and within science across different grades and measurement points were high (unadjusted: mean $r_{\text{math}} = .92$, mean $r_{\text{science}} = .85$; adjusted: mean $r_{\text{math}} = .94$, mean $r_{\text{science}} = .90$). The correlations between mathematics and science within the same year same grade studies (unadjusted: mean $r = .93$; adjusted: mean $r = .95$) and across years and grades were high as well (uncorrected: mean $r = .82$; corrected: mean $r = .86$).

The correlations in *PISA* within reading, within mathematics and within science tests across the two different measurement points were very high: unadjusted $r_{\text{read}} = .92$ (adjusted: $r_{\text{read}} = .94$), unadjusted $r_{\text{math}} = .96$ (adjusted: $r_{\text{math}} = .97$) and unadjusted $r_{\text{science}} = .92$ (adjusted: $r_{\text{science}} = .95$), but the correlations among the three or four tests within studies in the same years (unadjusted mean $r = .96$, adjusted: $r = .98$) and across years and grades were similarly high (unadjusted mean $r = .90$, adjusted: $r = .94$).

These correlation patterns provide some evidence for the convergent validity of the content-specific scales at the macro-social level (the correlations between the same scales across different grades and measurement points were higher than the correlations between



DOI: 10.1002/per

Table 1. Correlations between different scales of student assessment studies across nations, unadjusted (on the top right) and adjusted (at the bottom left) (see methodological problems section for the corrections)

| | R9 | R14 | TM4 | TS4 | TM8 | TS8 | TM8 | TS8 | TM4 | TS4 | TM8 | TS8 | PR0 | PM0 | PS0 | PR3 | PM3 | PS3 | PP3 | PR1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| IEA-Read-91-9 | *.91* | .83 | .11 | .49 | .22 | .28 | .51 | .57 | −.11 | .26 | .46 | .63 | .84 | .69 | .80 | .78 | .64 | .76 | .66 | .24 |
| IEA-Read-91-14 | .89 | *.91* | .32 | .59 | .23 | .22 | .79 | .87 | .76 | .92 | .78 | .88 | .55 | .56 | .59 | .57 | .50 | .59 | .57 | .31 |
| TIMSS-Math-95-4 | .53 | .49 | *.97* | .87 | .95 | .88 | .94 | .82 | .89 | .85 | .95 | .86 | .43 | .57 | .64 | .49 | .66 | .75 | .69 | .77 |
| TIIMSS-Scien-95-4 | .79 | .64 | .88 | *.96* | .79 | .85 | .77 | .86 | .63 | .78 | .72 | .81 | .73 | .75 | .83 | .70 | .74 | .77 | .75 | .91 |
| TIMSS-Math-95-8 | .61 | .33 | .94 | .86 | *.92* | .89 | .94 | .84 | .76 | .74 | .93 | .84 | .28 | .52 | .53 | .27 | .58 | .69 | .64 | .70 |
| TIIMSS-Scien-95-8 | .71 | .41 | .88 | .91 | .96 | *.91* | .89 | .94 | .47 | .60 | .85 | .90 | .15 | .22 | .42 | .19 | .33 | .51 | .35 | .69 |
| TIMSS-Math-99-8 | .82 | .65 | .94 | .78 | .95 | .91 | *.98* | .96 | .89 | .86 | .97 | .94 | .75 | .86 | .83 | .73 | .87 | .87 | .87 | .80 |
| TIIMSS-Scien-99-8 | .93 | .70 | .82 | .87 | .90 | .95 | .96 | *.98* | .86 | .91 | .92 | .96 | .81 | .87 | .86 | .80 | .90 | .91 | .91 | .87 |
| TIMSS-Math-03-4 | .19 | .69 | .96 | .73 | .88 | .70 | .93 | .92 | *.98* | .97 | .93 | .90 | −.07 | .31 | .20 | .68 | .82 | .88 | .84 | .87 |
| TIIMSS-Scien-03-4 | .38 | .77 | .89 | .88 | .84 | .81 | .89 | .95 | .98 | *.99* | .87 | .93 | −.01 | .24 | .18 | .79 | .83 | .89 | .87 | .95 |
| TIMSS-Math-03-8 | .67 | .80 | .97 | .74 | .92 | .87 | .98 | .95 | .95 | .92 | *.98* | .95 | .73 | .84 | .80 | .73 | .87 | .89 | .87 | .73 |
| TIIMSS-Scien-03-8 | .82 | .86 | .80 | .80 | .86 | .89 | .94 | .97 | .93 | .96 | .96 | *.98* | .87 | .93 | .90 | .88 | .95 | .97 | .96 | .78 |
| PISA-Reading-00 | .89 | .71 | .47 | .76 | .31 | .25 | .82 | .90 | .35 | .58 | .77 | .89 | *.97* | .95 | .97 | .92 | .90 | .89 | .89 | .63 |
| PISA-Math-00 | .82 | .58 | .59 | .76 | .49 | .29 | .90 | .93 | .57 | .52 | .84 | .92 | .97 | *.97* | .97 | .89 | .96 | .93 | .96 | .62 |
| PISA-Science-00 | .89 | .71 | .62 | .81 | .48 | .40 | .87 | .92 | .54 | .60 | .81 | .90 | .99 | .98 | *.97* | .87 | .90 | .92 | .90 | .69 |
| PISA-Reading-03 | .90 | .71 | .61 | .80 | .38 | .31 | .85 | .88 | .76 | .84 | .78 | .86 | .94 | .93 | .92 | *.95* | .94 | .93 | .94 | .55 |
| PISA-Math-03 | .83 | .61 | .75 | .80 | .58 | .39 | .93 | .93 | .89 | .86 | .88 | .91 | .94 | .97 | .94 | .97 | *.97* | .97 | .99 | .52 |
| PISA-Science-03 | .91 | .72 | .82 | .83 | .64 | .49 | .92 | .93 | .94 | .91 | .91 | .93 | .94 | .96 | .95 | .97 | .98 | *.95* | .97 | .69 |
| PISA-Problems-03 | .84 | .66 | .76 | .80 | .61 | .39 | .93 | .94 | .91 | .90 | .90 | .93 | .93 | .97 | .94 | .97 | .99 | .98 | *.97* | .59 |
| PIRLS-01 | .16 | .18 | .83 | .96 | .85 | .84 | .83 | .90 | .91 | .97 | .81 | .83 | .88 | .83 | .88 | .79 | .72 | .82 | .74 | *.96* |
| *N* | 26 | 30 | 25 | 25 | 39 | 39 | 37 | 37 | 24 | 24 | 44 | 44 | 41 | 41 | 41 | 40 | 40 | 40 | 40 | 33 |
| g-loading, uncorr. | .95 | .95 | .97 | .98 | .94 | .94 | .98 | .98 | .97 | .98 | .97 | .99 | .99 | .99 | .99 | .99 | 1 | 1 | 1 | .97 |
| g-loading, corrected | .99 | .96 | .98 | .99 | .97 | .97 | .99 | .99 | .98 | .99 | .98 | .98 | .99 | 1 | .99 | .99 | 1 | 1 | 1 | .98 |

*Note*: Correlations between unadjusted and adjusted scales in the diagonal in italics; *g*-loading: loading on first unrotated factor (including IQ-tests), *g*-loading of IQ (Lynn & Vanhanen): $\lambda_{uncorr} = .96$ and $\lambda_{corr} = .96$, variance explained by first unrotated factor unadjusted 93.7% and adjusted 94.8%, $RMSEA_{uncorr} = .04$ and $RMSEA_{corr} = .03$. Adjusted data means not correction for attenuation, but corrected raw data (e.g. for low school attendance rates).



DOI: 10.1002/per

different scales across different grades and measurement points). However, the correlations between different scales within the same grades or measurement points were always higher than the above correlations. Thus, the discriminant validity of the specific scales as measures of nations is low.

Factor analyses were done with MPLUS statistical software using Full-Information-Maximum-Likelihood (FIML; Raykov, 2005). This kind of analysis allows for the use of all data (no listwise deletion of a country and all its information if one observation in one variable is missing). In a factor analysis the first unrotated factor (*g*-factor) explained 94% (unadjusted) or 95% (adjusted) of the variance of the 20 student assessment scales and the intelligence test collection of Lynn and Vanhanen (see Table 1 and Figure 1 a,b). Thus, cognitive ability differences across nations are by and large unidimensional. The low

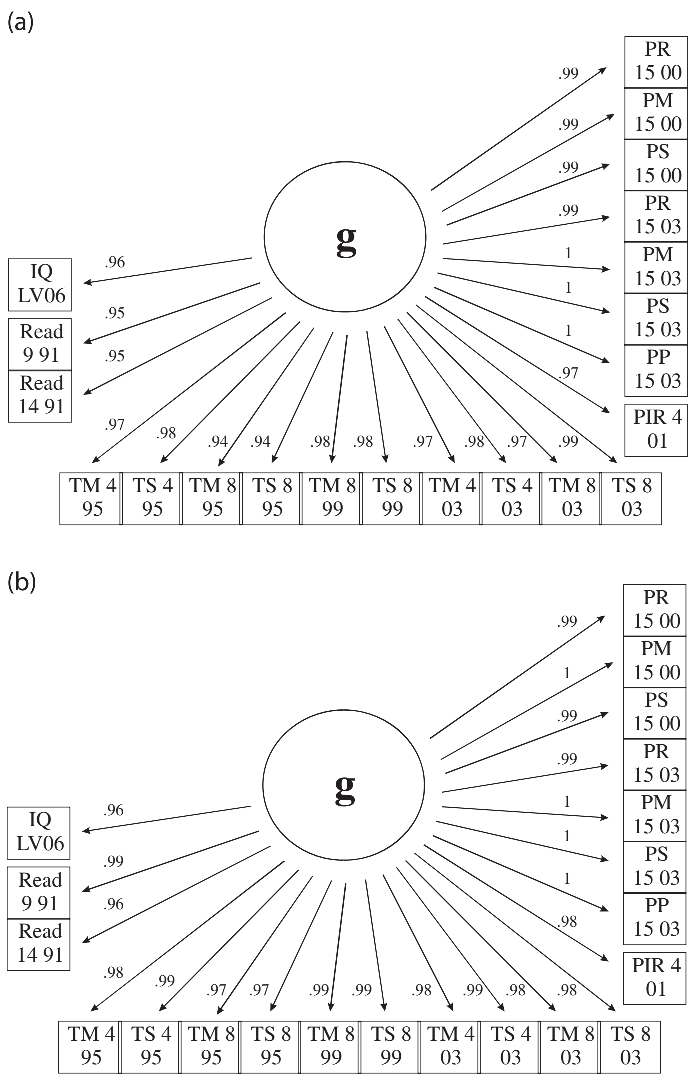


Figure 1. (a) *g*-factor of cognitive abilities at national level (unadjusted data) and (b) *g*-factor of cognitive abilities at national level (adjusted data, no correction for attenuation).



DOI: 10.1002/per

covariance coverages in FIML were problematic; in this situation FIML seems to increase the correlations slightly. But earlier studies with listwise deletion (14 countries and only scales from PISA 2000 and 2003) gave similar results (93% of variance was explained by the first unrotated factor; Rindermann, 2006a).

## Correlations among same scales of different student assessment approaches

IEA-Reading, PISA and PIRLS measure reading, and TIMSS and PISA measure mathematics and science. Correlations between scales with the same content should be higher than those between scales with different content.

*IEA-Reading* and *PISA-Reading* (15-year-old students) correlated using unadjusted data and in the mean with $r_{read} = .69$ (adjusted: $r_{read} = .80$); the correlations of similar age groups (unadjusted IEA-Read-14 with PISA, $r_{read} = .56$, adjusted: $r_{read} = .71$) were lower than the correlations of different age groups (unadjusted IEA-Read-9, $r_{read} = .81$, adjusted: $r_{read} = .90$), and the correlations with other PISA tests were nearly the same (unadjusted: $r = .64$, adjusted: $r = .76$).

*IEA-Reading* and *PIRLS* showed a very low correlation (unadjusted: $r_{read} = .28$, adjusted: $r_{read} = .17$), but the country sample was small ($N = 16$).

*TIMSS* and *PISA* mathematics correlated with unadjusted mean $r_{math} = .69$ (adjusted: $r_{math} = .74$); TIMSS and PISA science correlated with unadjusted mean $r_{science} = .72$ (adjusted: $r_{science} = .77$); and correlations between mathematics and science tests were similar (unadjusted: $r = .69$, adjusted: $r = .74$). In studies from similar years (TIMSS 1999 and PISA 2000), the pattern was about the same unadjusted $r_{math} = .86$ (adjusted: $r_{math} = .90$), unadjusted $r_{science} = .86$ (adjusted: $r_{science} = .92$), and unadjusted correlation across tests was $r = .85$ (adjusted: $r = .90$).

*PIRLS* and *PISA-Reading* correlated with unadjusted (mean) $r_{read} = .59$ (adjusted: $r_{read} = .84$); the correlations with other PISA scales were similar (unadjusted: $r = .62$, adjusted: $r = .80$).

To summarise, across IEA and OECD studies the discriminant validity of the specific tests as measures of nations was low. The correlations of *IEA-Reading* with other studies were in general low. In this oldest study there were probably some more serious unknown problems. Correlations within the studies (e.g. grades, measurement points, study organiser IEA or OECD) were higher than correlations across studies. But some statistical support for *content-specific differences* is given by TIMSS and PISA. For example, in Israel, there were higher values in reading than in mathematics and science (PISA 2000: 452 vs. 433 and 434); the same verbal pattern is shown in intelligence tests for Jews of other countries by Herrnstein and Murray (1994, p. 275), Lynn (2004), and Weiss (2000). Countries in East Asia (Japan, South Korea, Hong Kong, Macau, Singapore, Taiwan) were better in mathematics and science (mean of PISA 2000 and 2003: 541) than in reading (mean of PISA 2000 and 2003: 514). This corresponds to intelligence test results with Asian samples in the USA (Oesterdiekhoff, 2000, p. 103, verbal IQ 97 vs. nonverbal 110); modest quantity of books at home and school (and if yes, then science books); high interest for science, instruction in a foreign language (Hong Kong, Macau, Singapore); and silence in the classroom (Hesse, 2007, p. 255). Asian pupils and their parents seem not to put emphasis on reading, and teachers and their pupils do not appear to emphasise speaking. The pattern of results (lower in reading/language/verbal competence) corresponds to this.



DOI: 10.1002/per

### Correlations among sum scales of different student assessment studies

In the *IEA-Reading-Study 1991*, the correlation between reading levels of the 9- and 14-year old pupils was high (unadjusted: $r = .83$, adjusted: $r = .89$; see Table 1). Correlations with other studies were rather low (unadjusted: $r = .66$, adjusted: $r = .72$; see Table 2), with the highest being to the intelligence test collection of Lynn and Vanhanen (unadjusted: $r = .88$, adjusted: $r = .90$). The *IQs* of the Lynn and Vanhanen study correlated highly with all other study values; one reason is the larger sample and distribution of abilities (see Table 2). After adjustment, *PIRLS* shows consistently close relationships to other studies, except for the IEA-Reading study.

### Correlations among age- and grade-oriented studies

The correlations among the grade-oriented IEA-Studies (TIMSS and PIRLS) and the age-oriented OECD-Studies were high (unadjusted: $r = .87$, adjusted: $r = .91$, $N = 39$; see Table 3). TIMSS and PISA correlated with $r = .82$ (adjusted: $r = .83$, $N = 38$). Aggregation across grades and measurement points increased the correlations. Ability levels were more reliably measured. Even the relationships of IEA-Reading are not problematic any more.

### Correlations among student assessment studies and intelligence test studies

Correlations among different student assessments and intelligence test studies and among the sums of student assessment studies and intelligence test studies were high (see Table 3): Unadjusted TIMSS scores correlated with national intelligence levels $r = .89$ (adjusted $r = .88$, $N = 63$), as did grade-oriented studies with national intelligence (unadjusted TIMSS and PIRLS sum $r = .89$, adjusted $r = .89$, $N = 65$). Likewise, unadjusted PISA sum correlated with national intelligence $r = .88$ (adusted $r = .84$, $N = 46/47$), and the unadjusted sum of student assessment studies correlated with intelligence tests $r = .86$ (adjusted $r = .85$, $N = 76/77$). Adjustments and aggregations did not always lead to higher correlations. Analyses with aggregated data covered wider samples with more developing countries, but had less valid and stable data.

Lynn and Vanhanen (2006) measured intelligence levels in $N = 113$ countries and estimated them for another $N = 79$ countries. In comparison, *measured IQ* correlated with unadjusted sum of student assessment studies $r = .87$ (adjusted $r = .87$, $N = 63$) and *estimated IQ* correlated $r = .74$ (adjusted $r = .62$, $N = 13$). Thus, their estimation process seems to have been viable.

The correlations among the different cognitive ability scales remained stable when GNP was partialled: The unadjusted full correlation between grade- and age-oriented studies was $r = .87$, while the partial correlation was $r_p = .85$; the adjusted correlations were $r = .90$, $r_p = .84$, $N = 39$. The unadjusted full correlation between sum of student assessment studies and intelligence test studies was $r = .86$, and the partial correlation was $r_p = .78$ (adjusted $r = .85$, $r_p = .73$; $N = 76/77$). Thus, these relationships could not be attributed directly to national wealth.

### Correlations with further cognitive ability measures and attributes of societies

Not surprisingly, the correlations of the analysed IEA-, OECD- and intelligence-test studies were highest with other cognitive ability and education measures (see Table 4):



DOI: 10.1002/per

Table 2. Correlations between sum values of different cognitive ability studies across nations, unadjusted (on the top right) and adjusted (at the bottom left) (see methodological problems section for the corrections)

| | IEA-R-91 | TI-95-4 | TI-95-8 | TI-99-8 | TI-03-4 | TI-03-8 | PISA-00 | PISA-03 | PIRLS-01 | IQ-LV-06 |
|---|---|---|---|---|---|---|---|---|---|---|
| IEA-Read-91 | *.91* | .46 | .28 | .76 | .79 | .82 | .84 | .80 | .30 | .88 |
| TIMSS-95-4 | .70 | *.96* | .92 | .92 | .86 | .91 | .70 | .75 | .87 | .82 |
| TIMSS-95-8 | .46 | .94 | *.91* | .94 | .70 | .92 | .42 | .53 | .72 | .85 |
| TIMSS-99-8 | .79 | .93 | .95 | *.98* | .90 | .97 | .86 | .89 | .85 | .89 |
| TIMSS-03-4 | .84 | .94 | .85 | .94 | *.98* | .93 | .17 | .85 | .93 | .87 |
| TIMSS-03-8 | .87 | .91 | .91 | .98 | .96 | *.98* | .86 | .92 | .76 | .92 |
| PISA-00 | .87 | .72 | .41 | .91 | .57 | .87 | *.97* | .95 | .65 | .86 |
| PISA-03 | .87 | .83 | .53 | .94 | .91 | .91 | .96 | *.96* | .61 | .87 |
| PIRLS-01 | .19 | .92 | .85 | .88 | .95 | .83 | .87 | .78 | *.96* | .78 |
| IQ-LV-06 | .90 | .84 | .85 | .91 | .88 | .92 | .84 | .84 | .81 | *.98* |
| *N* | 31 | 25 | 39 | 37 | 24 | 44 | 41 | 40 | 33 | 192/193 |

*Note*: Correlations between unadjusted and adjusted scales in the diagonal in italics; in IQ-LV-06 adjusted is Liechtenstein included (estimated by H.R.).



DOI: 10.1002/per

Table 3. Correlations between different cognitive ability study approaches—national data level, unadjusted (on the top right) and adjusted (at the bottom left)

| | IEA-R 91 | TIMSS | Grade | Age | Student | IQ-LV-06 | All |
|---|---|---|---|---|---|---|---|
| IEA-Reading-91 | *.91* | .77 | .84 | .81 | .95 | .88 | .94 |
| TIMSS (95-03) | .82 | *.97* | .99 | .82 | .98 | .89 | .97 |
| Grade (TIMSS + PIRLS) | .86 | .99 | *.98* | .87 | .99 | .89 | .98 |
| Age (PISA, 00-03) | .87 | .83 | .90 | *.97* | .97 | .88 | .97 |
| Student assessment sum | .95 | .98 | .99 | .99 | *.97* | .86 | .98 |
| IQ-LV-06 | .90 | .88 | .89 | .84 | .85 | *.98* | .99 |
| All cognitive ability sum | .95 | .97 | .99 | .97 | .98 | .99 | *.98* |
| *N* | 31 | 63 | 65 | 48 | 78 | 192/193 | 194 |

*Note*: Correlations between unadjusted and adjusted scales in the diagonal in italics; in Grade is TIMSS included, in student assessment sum are IEA-R 91, grade and age included, in all cognitive ability sum are student assessment sum and IQ-LV-06 included, in IQ-LV-06 adjusted is Liechtenstein included (estimated by H.R.).

Unadjusted cognitive ability sum with Adult-Literacy: $r = .69$, $N = 20$; unadjusted cognitive ability sum with educational level $r = .76$, $N = 173$ (adjusted: with Adult-Literacy $r = .73$ and with educational level $r = .78$). Correlations with wealth and democracy were lower (with unadjusted cognitive ability sum: $r = .61$, $N = 185$, and $r = .53$, $N = 183$; adjusted: $r = .64$ and $r = .56$). This correlation pattern supports the validity of the national cognitive ability measures. The correlation between unadjusted cognitive ability and educational level (see too Barber, 2005) remained high even after partialling GNP: full $r = .76$, partial $r_p = .61$ (adjusted: $r = .78$, $r_p = .63$; $N = 173$).

Correlations involving the methodically criticised CFT-study by Buj (1981) were also high, except for PISA ($r = -.01$, corrected: $r = -.01$, $N = 21$, Ghana not included). For example, the correlation with the unadjusted student assessment sum was $r = .71$ (adjusted: $r = .70$, $N = 22$, Ghana included; but Ghana excluded and only European countries: $r_{uncorr} = -.04$ and $r_{corr} = -.08$, $N = 21$). The Buj study showed high correlations only by including a country from Africa. The study *IAEP-II,* which has been challenged for low representativeness of pupils, showed even higher and more stable correlations with all variants of cognitive ability measurements at the macro-social level (for example, the correlation with unadjusted cognitive ability sum was $r = .88$, $N = 19$; adjusted: $r = .89$). The results for the *old student assessment studies* were similar (the correlation with unadjusted cognitive ability sum was $r = .95$, $N = 19$; adjusted: $r = .95$). Even very old data about the '*intellectual' order of immigrants* 1921 to the USA (mainly from European countries) showed substantial correlations (for example, with unadjusted cognitive ability sum: rank $r = .44$, $N = 14$; adjusted rank $r = .53$).

Adjusted correlations were somewhat higher than unadjusted correlations (see Table 4). This was true for the OECD-Adult-Literacy-Study (mean unadjusted: $r = .66$; adjusted: $r = .79$), for the educational level of adults (mean unadjusted: $r = .74$; adjusted: $r = .75$), for GNP (unadjusted: $r = .62$; adjusted: $r = .65$), and for democracy (mean unadjusted: $r = .57$; adjusted: $r = .58$). These differences support the plausibility of the adjustments.

Factor analyses with IQ-tests, grade-level oriented studies and age-level oriented studies together with different educational and national variables (GNP, political attributes) resulted in three-dimensional solutions. The dimensions were cognitive abilities, wealth, and political attributes or cognitive abilities and educational variables together, wealth, and



DOI: 10.1002/per

Table 4. Correlations between cognitive abilities and society attributes across nations

| | OECD-Adult-Literacy 90s | Intelligence tests of Buj 70s | Intellectual order of immigrants 1921 USA | IAEP-II math & science (9 and 13 years old) 1991 | Student assessment studies of 60s and 70s | Educational level adults end of 20th century | Gross national product 1998 | Democracy 1950–2004 | *N* |
|---|---|---|---|---|---|---|---|---|---|
| Unadjusted data | | | | | | | | | |
| Grade (TIMSS + PIRLS) | .61 | .73 | .23 | .84 | .84 | .72 | .57 | .55 | 65 |
| Age (PISA, 00-03) | .71 | −.01 | .73 | .74 | .70 | .75 | .72 | .67 | 48 |
| Student assessment sum | .66 | .71 | .49 | .87 | .86 | .73 | .60 | .58 | 78 |
| IQ-LV-06 | .63 | .73 | .25 | .89 | .95 | .75 | .62 | .52 | 193 |
| All cognitive ability sum | .69 | .73 | .44 | .88 | .95 | .76 | .61 | .53 | 194 |
| *N* | 20 | 22 | 14 | 19 | 19 | 173 | 185 | 183 | |
| Adjusted data | | | | | | | | | |
| Grade (TIMSS + PIRLS) | .70 | .69 | .21 | .83 | .84 | .74 | .61 | .57 | 65 |
| Age (PISA, 00-03) | .75 | −.01 | .71 | .80 | .67 | .73 | .73 | .65 | 48 |
| Student assessment sum | .70 | .70 | .53 | .89 | .85 | .75 | .64 | .59 | 78 |
| IQ-LV-06 | .63 | .73 | .25 | .90 | .95 | .77 | .63 | .55 | 193 |
| All cognitive ability sum | .73 | .72 | .53 | .89 | .95 | .78 | .63 | .56 | 194 |
| *N* | 20 | 22 | 14 | 19 | 19 | 173 | 185 | 183 | |

*Note*: In IQ-LV-06 06 is test data of Buj included, intellectual order of immigrants 1921 USA: rank correlation; adjusted data means not correction for attenuation, but corrected raw data (e.g. for low school attendance rates), IAEP-II data also corrected.

political attributes (details can be obtained from the author). In contrast, cognitive ability differences across nations were by and large unidimensional.

## DISCUSSION

The cross-national correlations between different scales, between different studies (e.g. grades/ages, measurement points, used scales) and between different approaches (e.g. IEA vs. OECD, grade-level vs. age-level, student assessment vs. intelligence tests) were generally high. Factor analyses supported a strong *g*-factor. Different scales of student assessment studies and different cognitive test approaches appear to have measured essentially the same construct, namely general national cognitive ability.

Similarly high correlations were found in analyses at the German state level (16 Deutsche Bundesländer; see Rindermann, 2006a, p. 82). The high correlations among different studies and years underline the quality of the studies, too (see similar results using only TIMSS 2003, Lynn & Mikk, 2007). There were serious problems within all studies, but they seem to be less serious than many critics have assumed, at least as far as the highly robust national or state data are concerned. Aggregation and corrections help to increase the reliability and validity of international cognitive ability comparisons.

Sum values are available for student assessment studies in 78 countries (see Appendix); the sums for all cognitive abilities are available for 194 countries (see Appendix). Of these 194 countries, there are measured data for 128 countries (66%), with estimated data for 66 countries (34%). These highly consistent cross-country differences are graphically depicted in Figure 2a–c.

But *why* are these national abilitiy differences so homogenous? There are probably several reasons, which are now discussed.

### Possible causes of homogeneity

*Analysis of content: items of different scales in student assessment studies are similar*
Within the PISA studies the tasks on different scales (items of verbal, mathematic, science and problem solving content) are very similar. In all scales pupils are presented information in the form of (long) texts, graphs and tables, which are to be read, understood and used for item solution. Many tasks are rather difficult to ascribe to verbal or mathematics or science or problem solving content, because all the scales presented content and demands for skills related to those in other scales (e.g. geographic and biological topics are inherent in the Sahara reading task; figural information similar to reading, science and problem solving tasks is inherent in the race mathematics task). The specific knowledge requirements are rather small; the majority of problems can be solved with the information provided in the tasks (Rindermann, 2006a). Literacy items tend to be longer versions of intelligence test tasks (see too Prais, 2003). In many intelligence test items like those in the Wechsler Adult Intelligence Scale or CAT, the knowledge requirement is clearer but is not useful for measuring intelligence as thinking ability.

Tasks in PIRLS are constructed in the same way as in PISA. But TIMSS tasks are different: they are shorter and they have a clearer knowledge content that make assignment to a particular subject area easier.

The use of multiple choice answer formats is also an important aspect of task analysis. Backward strategies make it possible to reduce the probability of wrong answers and



DOI: 10.1002/per

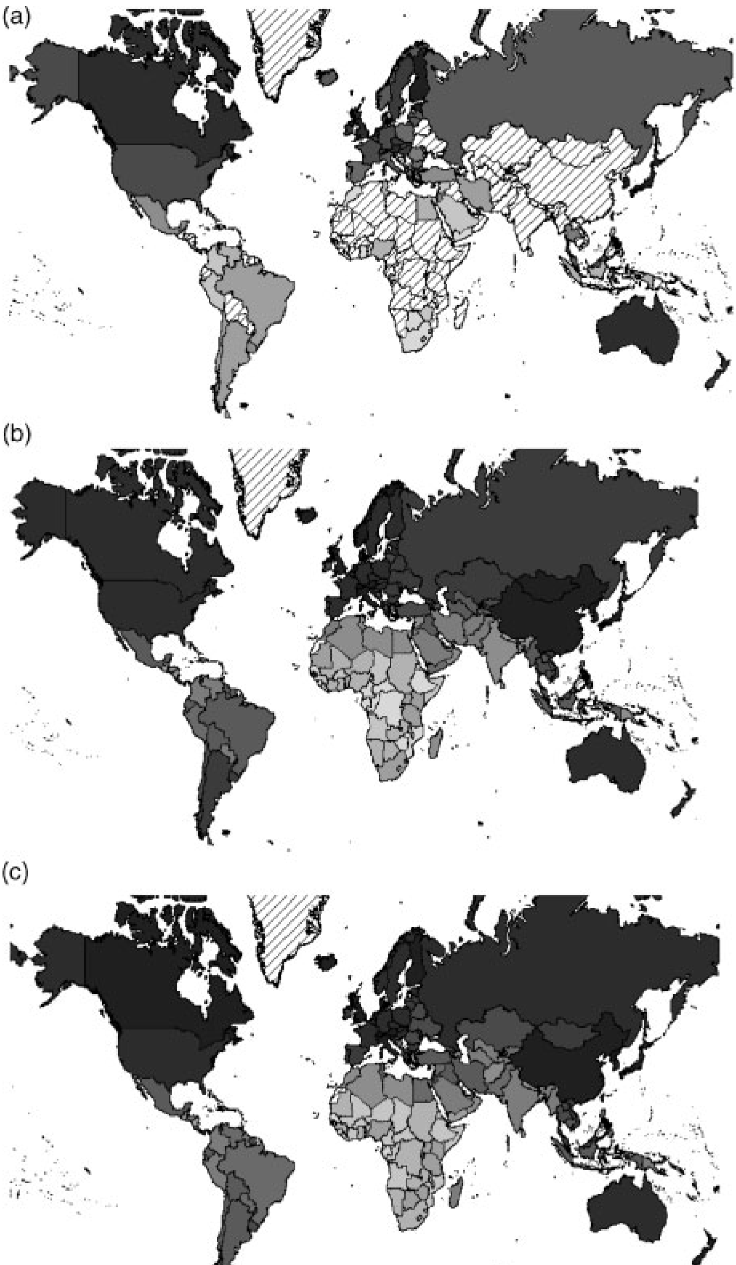


Figure 2. (a) Sum value of all student assessment studies ($N = 78$, unadjusted data, darker means higher, with hachures: no data; maps done with SAS), (b) distribution of measured and estimated intelligence test results from Lynn and Vanhanden ($N = 192$, unadjusted data, darker means higher, no values for West Sahara and Greenland) and (c) sum value of all (complex) cognitive ability studies ($N = 194$, adjusted data, darker means higher, no data for West Sahara and Greenland).



DOI: 10.1002/per

increase the influence of intelligence on test results. Additionally, minor errors in tasks and vagueness in wording questions (Meyerhöfer, 2006) make correcting interpretation necessary. This may increase the influence of intelligence on test scores. Systematic task analyses have to research the impact of these factors.

### *Analysis of demands: cognitive processes necessary for solving similar and different cognitive ability tasks are similar*

Tasks of different scales in PISA are similar. And in both student assessment tests and intelligence tests, thinking and knowledge tasks are mixed. There is no categorical difference between student achievement and intelligence assessment (e.g. Ceci, 1991). More convincing would be a differentiation between *thinking* and *knowledge* tasks.

Even for different tasks, the cognitive processes necessary for solving them are similar: (1) Information given in the task has to be found, stored and compared with information retained in memory. (2) Information has to be structured and understood. (3) Reasoning processes are necessary. (4) Common grounds and differences have to be determined. (5) Abstract thinking (e.g. categorisation) is necessary. (6) Mental speed, concentration, working memory, time management, motivation, low test anxiety, and cognitive and test-taking routines are helpful. (7) The time pressure (especially in TIMSS; Woschek, 2005) could increase the demand for general intelligence, too.

The theory behind student assessment (particularly for literacy) relies on complex assumptions (e.g. modelling; Kintsch, 1998; Murray et al., 1997). Some scepticism about the empirical validity of these assumptions, and in particular about the applicability and fit of this student assessment theory to given tasks, is advisable (Meyerhöfer, 2006). It is not unreasonable to hypothesise that the cognitive demands and processes involved in solving student assessment test tasks are similar to the ones involved in solving classical intelligence test tasks (e.g. Raven Matrices; Carpenter, Just, & Shell, 1990).

### *Analysis of cognitive development: causes for high abilities are similar and nested in reciprocal causation at the individual level*

Family background and school education simultaneously affect intelligence, knowledge and performance on a variety of cognitive tests. Beneficial environments have positive impacts on cognitive and motivational development, which results in high cognitive ability. Obviously, special knowledge such as about foreign languages and specific mathematical formulas (e.g. Pythagorean theorem) is not generally displayed without special instruction. But the quality of environment of children is normally relatively homogenous: parents who speak in their mother tongue in a grammatically correct manner probably also use a large vocabulary, probably also help their children with school problems, also explain their educational rules, also send their children to better schools with better classmates and more demanding instruction (e.g. Rindermann, 2007). Intelligence and knowledge are both a result of successful education. And they develop with reciprocal causation (see Maas, Dolan, Grasman, Wicherts, Huizenga, & Raijmakers, 2006), with positive beneficial interactions between intelligence, knowledge, environment and experiences (investment theory, positive reinforcement by environment, stimulation etc.).

In a famous and careful study, Hart and Risley (1995) investigated this homogenous influence of parents on different aspects of cognitive development (see Figure 3). Education and language used by parents influenced both intelligence and verbal development in a similar way, and were more important than socio-economic status.



DOI: 10.1002/per

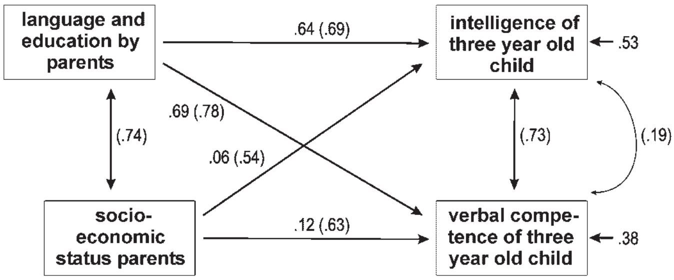


Figure 3. Determinants of intelligence and verbal development of 3-year-old children (standardised path coefficients, correlations in parentheses, correlated error between intelligence and verbal competence, error terms as not explained variance on the right, $N = 42$, reanalysis by the author, data from Hart & Risley, 1995).

Further educational research underscores the positive impact of education on school performance and cognitive development (e.g. Dornbusch et al., 1987).

Additionally, basic cognitive processes—especially mental speed—are helpful for intelligence and students performance at school (Luo, Thompson & Detterman, 2003; Rindermann & Neubauer, 2004). According to Cattell's (1987/1971) investment theory of cognitive development, fluid intelligence, like (very basic) mental speed, is necessary for the development of crystallised intelligence, including knowledge.

### *Analysis of cognitive development: causes for high abilities are similar at the macro-social level*

Factors contributing to high correlations at the individual level (similarity of items, similarity of cognitive demands, similarity of developmental factors in reciprocal causation) are also responsible for the high correlations at the national level, but they are complemented by higher level factors exclusive to the national level.[2] Again, conditions furthering thinking abilities are similar to the ones furthering knowledge in relevant domains: competent and extensive education in families and in different kinds of schools (from kindergarten to university); a high formal education rate of the population (secondary and university education); wealth (nutrition, health, education system); high esteem for education, achievement, knowledge, thinking and rationality in families and in an educational system established by a culture that approves of achievement, knowledge, thinking and rationality, especially thinking for oneself oriented towards rationality ('sapere aude'). Social, economic and cultural liberty framed by the rule of law, as well as a peaceful, non-violent interaction structure based on arguments, not on power or force, and a meritorious social organisation further education, thinking, the acquisition of knowledge and achievement. All these are aspects of a civil society and its norms on which nations may differ ('bürgerliche Gesellschaft'; first thorough description: Alberti, 2004/1441).

[2]Due to lack of space results of correlational and factor analytical studies at the individual level could not be presented. Manifest correlations between scales within student assessment studies were high at the individual level too (e.g. within TIMSS grade 8 1995 $r = .61$, within PISA 2000 $r = .65$ and within PISA 2003 $r = .71$, within PIRLS $r = .63$, within OECD-Adult-Literacy latent $r = .89$; Rindermann, 2006a). Correlations with intelligence tests were substantial (PISA with CAT $r = .47$, in PIRLS with CAT latent $r = .61$). Factor analysis in PISA (PISA 2000 and CAT scales) showed $g$-loadings that were always higher than specific factor loadings; the $g$-factor in a nested factor model explained 44% of the variance in PISA-scales, while the specific factors explained only 13% of the variance; test intelligence was not separated from PISA- or cognitive-ability-$g$ (Brunner, 2005).



DOI: 10.1002/per

More tangible is the positive influence of the intelligence of others and society within family, neighbourhood, peers, class, school, job, institutions (educational, health, economic, cultural systems) and society achieved through sophisticated communication, better instruction, more stimulating environment, modelling, critics and help: the intelligence of others creates intelligence.

*Genetic components*

At the *individual level* several studies have shown that inter-individual differences in performance on student assessment tests and intelligence tests share common genetic influences (twin studies of Bartels, Rietveld, Baal, & Boomsma, 2002; Kovas, Harlaar, Petrill, & Plomin, 2005; Wainwright, Wright, Geffen, Geffen, Luciano, & Martin, 2005). The specific genetic factors (e.g. which 'genes' and their functions in the production of proteins and neurological structure) are not yet known. In PISA the heritability probably is even higher because of the strong *g*-factor and the relatively small impact of knowledge.

Lynn, Rushton, Jensen, and others (Lynn & Vanhanen, 2002, 2006; Lynn, 2006; Rushton, 1997/1995; Rushton & Jensen, 2005; Templer & Arikawa, 2006) have tried to explain macro-social level differences by genetic theories. However, specific genes have not been found. Some hypotheses that do involve specific genes at the macro-social level have been presented in two papers by Bruce Lahn's research group in 2005 (Microcephalin and ASPM; Evans et al., 2005; Mekel-Bobrov et al., 2005). There are high correlations with the cognitive ability sum at national level (unadjusted: $r = .76$, adjusted: $r = .79$, $N = 24$; adjusted, GNP and educational level partialed out $r_p = .66$, $N = 24$). However, the status of these two hypotheses is currently controversial. Woods et al. (2006) and Rushton, Vernon and Bons (in press) found no relationship of Microcephalin and ASPM with brain size and intelligence at the level of the individual. Thus, the correlation between the frequency of Microcephalin and ASPM and cognitive ability at the macro-social level is presently unexplained. The quest for genes remains open.

## What term should be used for the competences measured by intelligence and student assessment tests?

Student assessment studies measure cognitive abilities. 'Student assessment' or 'student performance' themselves are not ability constructs. Intelligence is understood as the ability to think: it is the ability to solve new cognitive problems by thinking (without relying on pure recall of knowledge), to infer (to draw inductive and deductive-logical conclusions, reasoning), to think abstractly (to categorise, to sort out information, to process abstract information in the form of verbal and numerical symbols, in the form of abstract figures and in the form of general rules), and to understand and realise (to recognise and construct structures, relationships, contexts and meaning). Thinking ability includes the ability to change cognitive perspectives, and to planning and use foresight; in fact, learning depends on the use of these abilities (e.g. Rindermann, 2006a).

Pure knowledge is not included in this definition. But knowledge is always necessary for thinking, and intelligence tests, even figural tasks like Raven Matrices are never completely free from knowledge and experience (e.g. improvement of CFT-results by schooling, Stelzl, Merz, Remer, & Ehlers, 1995). Tests could be arranged according to their proximity to educational and school content ('school-distant and school-near cognitive ability') or their proximity to intelligence or to knowledge. This categorisation should be done within and across intelligence and student assessment tests, because in both types of tests,

 
DOI: 10.1002/per

intelligence and knowledge tasks are mixed. Raven Matrices and the figural tasks of CFT, CAT and BIS-T4 are more school-distant. Verbal and numerical tasks of usual intelligence tests, of PISA and PIRLS scales are less school-distant. Knowledge-based questions in intelligence tests like the WAIS and CAT or in the student assessment study TIMSS (e.g. necessary knowledge about foreign words, mathematical formula, natural science) are school-near. The sum value of different student assessment scales or of student assessment and intelligence test scales represents a combination of intelligence and knowledge. Knowledge itself has to be assessed normatively by its truthfulness and relevance. Pure knowledge questions (about true and important content) are only indirect measures of intelligence. A term that encompasses both intelligence and knowledge could be *general complex* (individual and national) *cognitive ability.*

The 'literacy' concept could be included in the historically older intelligence concept. Intelligence is a term frequently used in occidental history of thought. It was used in the Middle Ages (scholasticism), explicitly as a term for cognitive ability in the Age of Enlightenment (e.g. Condorcet, 1963/1794, p. 252, in French 'intelligence'). Since the 19th century, it has been the well-established term for the *ability to think* (e.g. Hegel, 1970/1822, p. 161, 'was Intelligenz, was Denken ist'). Of course, Enlightenment is more than intelligence. Enlightenment is rationality, reflection, intellectual autonomy and the ability to question, but intelligence is, besides knowledge, the central prerequisite for it.

In the student assessment reports, the term intelligence is never used, which is a strange phenomenon. Perhaps the term literacy was constructed to reduce political and legal problems of cognitive ability measurement in the USA. Perhaps it is an indication of doubts about thinking itself. Strangely and possibly for similar political reasons, the term Taiwan is not used in the student assessment reports, but Chinese Taipei is used. Perhaps 'literacy' is a kind of 'Chinese Taipei' for intelligence.

## Future research

Numerous important problems and questions remain open:

1. It seems possible to improve upon the *estimation process* of values for countries without intelligence test results (e.g. by regression). But the simple method of Lynn and Vanhanen (using the mean of neighbouring countries with similar racial/social composition) or the adjustment made here (deduction of five points), both have the advantage of avoiding the overestimation of the relationship with other society attributes. Something worth considering seems to be the aggregation formulas; it is possible to give bonus points for participation in student assessment studies. Similar the *adjustments* for age of pupils and low school attendance rates seem to be improvable.
2. The *representativeness of results* for pupils, for the youth of the same generation, for the entire youth, for the future adult population and for the population of a whole country depends on the representativeness of the samples, on the magnitude of a country, on the variation within countries (regional differences, e.g. China; class differences, e.g. India) and on youth–adult differences in the composition of societies. Large migrations or different birth rates can change them.
3. *Test-wiseness* (sophistication) or different traditions of test-use in different countries can bias the differences among countries (Wuttke, 2006). But the frequent use of tests could be itself a method to encourage education, learning and thinking.



DOI: 10.1002/per

4. If a more *specific (less* g*-oriented) measurement of cognitive abilities* (competences) is desired, the test scales should be constructed to be more knowledge-based. Correlations at the individual and national levels will remain because the conditions advancing thinking and knowledge in different domains are similar at both levels.
5. Apart from national cognitive ability means, the *variance* or *ability distribution* is of interest. For economic wealth, the 'Gini-coefficient' is used as a measure of differences (Rindermann, in press); similar coefficients can be developed for cognitive abilities, too (in the majority of student assessment studies this information is documented). For example, in Latin America there is a strong, but very small group with very high cognitive ability (represented by intellectuals like Mario Vargas Llosa, Gabriel García Márques, Pablo Neruda) in an environment of generally rather low cognitive ability; the same exists possibly for India. In the future, this may also be the case in countries with large immigration rates from countries with poor education.
6. Research done at the national level about the effectiveness of education or characteristics of societies (economic growth, wealth, democracy, liberty, peace, health, etc.) should be done with *sum values* for a large sample of countries. The use of single scales or single studies is scientifically inappropriate. This does not mean that well-founded content-specific hypotheses should not be developed and tested (exemplified by cultural, educational, social and genetic characteristics of Jews vs. East-Asians; e.g. Cochran, Hardy, & Harpending, 2006; Herrnstein & Murray, 1994; Weiss, 2000).
7. The *consequences of different national cognitive ability levels on the development of their societies* need to be researched (as has been done for economic questions; e.g. Lynn & Vanhanen, 2002; Weede & Kämpf, 2002). National cognitive ability levels are not only relevant for economic development, but also for other characteristics of societies such as democratisation, rule of law, technical and social modernisation (e.g. Oesterdiekhoff, 1992, 2000), and health (Kanazawa, 2006b).
8. The causes of *different national cognitive ability levels* have to be researched. Apart from educational determinants, genetic, economic, political and cultural factors are relevant, which themselves are not independent. There are likely to be complex reciprocal influences.
9. *Individual development of cognitive abilities depends on macro-social conditions of societies*. Birth in a modern society offering good and extensive education; with good nutrition and health systems; with high esteem for education, knowledge, thinking and rationality; and with a tradition of cultural and cognitive liberty makes the difference. However, some individualistic research traditions lead to a neglect of social determinants on and consequences of cognitive abilities.
10. As one important question, however, remains the *validity of the cognitive ability measures*. Are the test results of student assessment and intelligence test approaches valid indicators of the ability to think and to acquire and use true and valuable knowledge? There is much evidence that this is the case, including the high correlations between different student and intelligence test assessments, the high correlations of these test scores with indicators of education, and the correspondence of these scores with results of Piagetian tasks. Other approaches could be fruitful as well: the observation of intelligence and rationality in everyday life behaviour, as has been done in the Piagetian tradition (Oesterdiekhoff, 1992, 2000), but only rarely in the psychometric tradition (e.g. Gordon, 1997); for example, faith in witchcraft and rumour, or mistakes in daily behaviour as in traffic (e.g. Dagona, 1994). Measuring intelligence at the social and cultural level itself, e.g. by the availability of schools and



DOI: 10.1002/per

universities, by the estimation of the quality of newspapers and TV, by the effectiveness of government and the cognitive ability of political leaders (e.g. Simonton, 2006), by the quality of work of intellectuals, or by the rationality of the social organisation, the political system, the culture and the philosophies of life would be worth considering.

## General problems of present cognitive ability research

*Isolated research paradigms*

Cognitive ability research at the individual as well as the cross-cultural level takes place within different academic disciplines, university research fields, and education departments. Examples include ethnological or anthropological research (e.g. Lévy-Bruhl, 1966), cross-cultural developmental research in the tradiation of Piaget (e.g. Dasen, 1977), Marxist oriented cultural and cognitive developmental research (Lurija, 1976), combination of anthropological and Piagetian research (e.g. Hallpike, 1978), combination of Piagetian and sociological research (Oesterdiekhoff, 2000), psychometric intelligence research (e.g. Jensen, 1980, 1998), psychometric intelligence research in combination with an evolutionary approach (e.g. Lynn, 2006; Rushton, 1997), student assessment studies (PISA, TIMSS, PIRLS), and economic human capital research (e.g. Mankiw et al., 1992; Hanushek & Kimko, 2000).

These paradigms are highly isolated from another. In student assessment studies, the intelligence construct and authors such as Jensen, Lynn or Piaget are never mentioned. Different academic journals, conferences, societies, methods and technical terms hinder communication. Lack of knowledge, narrow ranges of interest and intellectual curiosity, language problems (e.g. English, French, German); and politically motivated attitudes against certain kinds of research make it easy to ignore other approaches. Research seems to be driven by tradition, personal background and disciplinary climate. In the measurement-oriented student assessment studies, more far-reaching research is sometimes neglected in favour of data collection and presentation; only reports are published.

Two exceptions exist in the form of analyses done with student assessment data by psychometricians at different data levels (Jensen, 1980; Herrnstein & Murray, 1994; Lynn & Vanhanen, 2002, 2006) and economic research using all kinds of cognitive ability research as indicators for human capital (e.g. Hanushek & Kimko, 2000; Weede & Kämpf, 2002; Carneiro & Heckman, 2003). Cognitive ability research could benefit from integrated research attempts, which consider different paradigms in empirical research.

*Lack of respect for alternative approaches*

The isolation of the different research traditions from one another is not only a chance result of different historical developments in science. Unfortunately, it is perpetuated by explicit disregard and disrespect of researchers from different research traditions. Papers submitted by researchers of other theoretical origin are regularly rejected, especially in educational journals. Critics of PISA are degraded publicly at the expense of their scientific reputations. In discussions about student assessment studies, an aggressive mood often dominates (e.g. description by Meyerhöfer, 2006). Critics of student assessment research provoke this mood, too (e.g. Freudenthal, 1975). Some terms provoke dispute, chiefly the terms *intelligence* or *genes* and particularly the combination of them, and the dispute seems endless because of sometimes uninformed usage of these concepts by persons involved in the debates. One factor that seems to contribute to this state of affairs is that the large-scale student assessment studies are done in cooperation between political institutions and



DOI: 10.1002/per

private companies; thus, names of constructs, research content, names of countries and even addresses of test companies ('Princeton') are influenced by political and economic interests (Flitner, 2006).

Cognitive ability research at the social and macro-social level has always been confronted with problems and pressures (see controversies around Jensen, Herrnstein & Murray, Rushton, Vanhanen; e.g. Segerstråle, 2000; Nyborg, 2003; Rindermann, 2006b). Intelligence and knowledge are important constructs, with perhaps the most impact on individual and national wealth, democracy and health in modern times. Science can grow only in a mood of rationality and with respect for the rules of thinking and argumentation, as well as respect for other researchers in a field.

## ACKNOWLEDGEMENTS

Christian Geiser gave me important advices for the analyses with MPLUS, Joachim Wuttke and James Flynn for the use of adjustment formulas. Thanks to Janel Schuh and Wendy Johnson for stylistic corrections. I have benefited from a lot of suggestions, help and criticicism received after talks, lectures, and distribution of preliminary drafts of this paper by colleagues and students from Germany, Austria and the USA and by three anonymous reviewers.

## REFERENCES

Alberti, L. B. (2004/1441). *The family in Renaissance Florence (I libri della famiglia)*. Long Grove: Waveland.

Ammermüller, A., & Pischke, J.-S. (2006). *Peer effects in European primary schools: Evidence from PIRLS*. Mannheim: Discussion paper no. 6–27.

Armor, D. J. (2003). *Maximizing intelligence*. New Brunswick: Transaction Publishers.

Badke-Schaub, P., & Strohschneider, S. (1998). Complex problem solving in the cultural context. *Le Travail Humain 61*, 1–28.

Barber, N. (2005). Educational and ecological correlates of IQ: A cross-national investigation. *Intelligence*, *33*, 273–284.

Barro, R. J., & Lee, J.-W. (1993). International comparisons of educational attainment. *Journal of Monetary Economics*, *32*, 363–394.

Barro, R. J., & Lee, J.-W. (2000). *Barro-Lee data set. International data on educational attainment: Updates and implications*. Boston: Harvard University. Retrieved November 18 2004 from www2.cid.harvard.edu/ciddata/barrolee/readme.htm

Bartels, M., Rietveld, M. J. H., Baal, G. C. M., & Boomsma, D. I. (2002). Heritability of educational achievement in 12-year olds and the overlap with cognitive ability. *Twin Research*, *5*, 544–553.

Beaton, A. E., Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Kelly, D. L., & Smith, T. (1996). *Mathematics achievement in the middle school years*. Chestnut Hill: TIMSS Study Center.

Beaton, A. E., Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., Smith, T., & Kelly, D. L. (1996). *Science achievement in the middle school years*. Chestnut Hill: TIMSS Study Center.

Berry, J. W., & Dasen, P. R. (Eds.). (1974). *Culture and cognition: Readings in cross-cultural psychology*. London: Methuen.

Bishop, J. H. (2004). *Drinking from the fountain of knowledge: Student incentive to study and learn*. Cornell: Center for Advanced Human Resource Studies (CAHRS).

Bracey, G. W. (2000). The TIMSS "Final Year" study and report: A critique. *Educational Researcher*, *29*, 4–10.

Brunner, M. (2005). *Mathematische Schülerleistung [student performance in mathematics]*. Berlin: Dissertation at HU Berlin.

Buj, V. (1981). Average IQ values in various European countries. *Personality and Individual Differences*, *2*, 168–169.

 
DOI: 10.1002/per

Carneiro, P., & Heckman, J. J. (2003). *Human capital policy.* In J. J. Heckman, & A. B. Krueger (Eds.), *Inequality in America: What role for human capital policy?* (pp. 77–239). Cambridge: MIT Press.
Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices test. *Psychological Review*, *97*, 404–431.
Cattell, R. B. (1987/1971). *Intelligence: Its structure, growth and action*. Amsterdam: Elsevier.
Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? *Developmental Psychology*, *27*, 703–722.
Cochran, G., Hardy, J., & Harpending, H. (2006). Natural history of Ashkenazi intelligence. *Journal of Biosocial Science*, *38*, 659–693.
Collani, Ev. (2001). OECD PISA—an example of stochastic illiteracy? *Economic Quality Control*, *16*, 227–253.
Condorcet, M.-J.-N. (1963/1794). *Entwurf einer historischen Darstellung der Fortschritte des menschlichen Geistes [Esquisse d'un tableau historique des progrès de l'esprit humain/Sketch for a historical picture of the progress of the human mind]*. Frankfurt: Europäische Verlagsanstalt.
Dagona, Z. K. (1994). Substance use and road traffic accidents among Nigerian commercial motor cyclists. *Ife PsychologIA*, *2*, 81–93.
Dasen, P. R. (Ed.). (1977). *Piagetian cross-cultural psychology.* NY: Halsted.
Dornbusch, S. M., Ritter, Ph. L., Leiderman, P. H., Roberts, D. F., & Fraleigh, M. J. (1987). The relation of parenting tyle to adolescent school performance. *Child Development*, *58*, 1244–1257.
DPK (Prenzel et al.) (2005). *PISA 2003*. Kiel. Retrieved November 4, 2005 from http://pisa.ipn.uni-kiel.de/PISA2003_E_Zusammenfassung.pdf
Elley, W. B. (1992). *How in the world do students read*? The Hague: IEA.
Evans, P. D., Gilbert, S. L., Mekel-Bobrov, N., Vallender, E. J., Anderson, J. R., & Vaez-Azizi, L. M., et al. (2005). Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. *Science*, *309*, 1717–1720.
Fertig, M. (2002). *Educational production, endogenous peer group formation and class composition. Evidence from the PISA 2000 study.* Essen: RWI Discussion Papers, 2.
Flitner, E. (2006). Pädagogische Wertschöpfung [educational value added]. In J. Oelkers, R. Casale, R. Horlacher, & S. Larcher Klee. (Eds.), *Rationalisierung und Bildung bei Max Weber [rationalization and education with Max Weber]* (pp. 245–266). Bad Heilbrunn: Klinkhardt.
Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, *101*, 171–191.
Freudenthal, H. (1975). Pupils' achievement internationally compared—The IEA. *Educational Studies in Mathematics*, *6*, 127–186.
Gaeth, F. (2005). *PISA. Eine statistisch-methodische evaluation [PISA. A statistical-methodical evaluation]*. Berlin: Dissertation at FU Berlin.
Gonzales, P. (2000). *Mathematics and science in the eigth grade*. Washington: National Center for Education Statistics.
Gordon, R. A. (1997). Everyday life as an intelligence test: Effects of intelligence and intelligence context. *Intelligence*, *24*, 203–320.
Gottfredson, L. (1997). Why *g* matters: The complexity of everyday life. *Intelligence*, *24*, 79–132.
Gottfredson, L., & Deary, I. J. (2004). Intelligence predicts health and longevity, but why? *Current Directions in Psychological Science*, *13*, 1–4.
Greenfield, P. M. (1966). On culture and conservation. In J. S. Bruner, R. R. Olver, & P. M. Greenfield (Eds.), *Studies in cognitive growth* (pp. 225–256). New York: Wiley.
Hallpike, Ch. (1978). *Foundations of primitive thought*. Oxford: University Press.
Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics*, *18*, 527–544.
Hanushek, E. A., & Kimko, D. D. (2000). Schooling, labor-force quality, and the growth of nations. *American Economic Review*, *90*, 1184–1208.
Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Paul Brookes.
Hegel, G. W. F. (1970/1822). *Vorlesungen über die Philosophie der Weltgeschichte. Band I: Die Vernunft in der Geschichte [lectures about philosophy in world history. Reason in history]*. Berlin: Akademie-Verlag.



DOI: 10.1002/per

Herrnstein, R. J., & Murray, C. H. (1994). *The bell curve*. New York: Free Press.
Hesse, H.-G. (2007). Lernen innerhalb und außerhalb der Schule aus interkultureller Perspektive [Learning inside and outside school from a cross-cultural perspective]. In G. Trommsdorff, & H.-J. Kornadt (Eds.), *Anwendungsfelder der Kulturvergleichenden Psychologie. Enzyklopädie der Psychologie. C/VII/3 [Cross-cultural psychology]* (pp. 187–278). Göttingen: Hogrefe.
Hunt E., & Wittmann, W. (in press). National intelligence and national prosperity. *Intelligence*, *35*. DOI: 10.1016/j.intell.2006.11.002
Jensen, A. R. (1980). *Bias in mental testing*. London: Methuen.
Jensen, A. R. (1998). *The g factor. The science of mental ability.* Westport: Praeger.
Kanazawa, S. (2006a). IQ and the wealth of states. *Intelligence*, *34*, 593–600.
Kanazawa, S. (2006b). Mind the gap... in intelligence: Re-examining the relationship between inequality and health. *British Journal of Health Psychology*, *11*, 623–632.
Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, UK: Cambridge University Press.
Kovas, Y., Harlaar, N., Petrill, S. A., & Plomin, R. (2005). 'Generalist genes' and mathematics in 7-year-old twins. *Intelligence*, *33*, 473–489.
Lapointe, A. E., Askew, J. M., & Mead, N. A. (1992). *The international assessment of educational progress: Learning science*. Princeton: Educational Testing Service.
Lapointe, A. E., Mead, N. A., & Askew, J. M. (1992). *The international assessment of educational progress: Learning mathematics*. Princeton: Educational Testing Service.
Lee, J.-W., & Barro, R. J. (1997). *Schooling quality in a cross-section of countries*. New York. Retrieved November 18, 2004 from www.nber.org/papers/W6198
Lévy-Bruhl, L. (1966). *Primitive mentality.* Boston: Beacon Press.
Loehlin, J. C. (2007). Book review of Lynn's (2006) "Race differences in intelligence". *Intelligence*, *35*, 93–94.
Luo, D., Thompson, L. A., & Detterman, D. K. (2003). The causal factor underlying the correlation between psychometric *g* and scholastic performance. *Intelligence*, *31*, 67–83.
Lurija, A. R., [Luria] (1976). *Cognitive development. Its cultural and social foundations*. Cambridge: Harvard University Press.
Lynn, R. (2004). The intelligence of American Jews. *Personality and Individual Differences*, *36*, 201–206.
Lynn, R. (2006). *Race differences in intelligence: An evolutionary analysis*. Atlanta: Washington Summit.
Lynn, R., & Mikk, J. (2007). National differences in intelligence and educational attainment. *Intelligence*, *35*, 115–121.
Lynn, R., & Vanhanen, T. (2002). *IQ and the wealth of nations*. Westport: Praeger.
Lynn, R., & Vanhanen, T. (2006). *IQ and global inequality.* Athens: Washington Summit.
Maas, H. L. J. v. d., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*, 842–861.
Mackintosh, N. J. (2007). Book review of Lynn's (2006) "Race differences in intelligence". *Intelligence*, *35*, 94–96.
Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A contribution to the empirics of economic growth. *Quarterly Journal of Economics*, *107*, 407–437.
Marshall, M. G., & Jaggers, K. (2000). *Polity IV Project. Political regime characteristics and transitions, 1800–1999*. University of Maryland. Retrieved October 5, 2004 from www.bsos.umd.edu/cidcm/inscr/polity
Martin, M. O., Mullis, I. V. S., Beaton, A. E., Gonzalez, E. J., Smith, T., & Kelly, D. L. (1997). *Science achievement in the primary school years*. Chestnut Hill: TIMSS Study Center.
Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 international science report*. Chestnut Hill: TIMSS & PIRLS International Study Center.
Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., Gregory, K. D., Smith, T. A., & Chrostowski, S. J., et al. (2000). *TIMSS 1999 international science report*. Chestnut Hill: TIMSS Study Center.
Meisenberg, G., Lawless, E., Lambert, E., & Newton, A. (2005). The Flynn effect in the Caribbean. *Mankind Quarterly*, *46*, 29–69.
Meisenberg, G., Lawless, E., Lambert, E., & Newton, A. (2006). Determinants of mental ability on a Caribbean island, and the mystery of the Flynn effect. *Mankind Quarterly*, *46*, 273–312.



DOI: 10.1002/per

Mekel-Bobrov, N., Gilbert, S. L., Evans, P. D., Vallender, E. J., Anderson, J. R., & Hudson, R. R., et al. (2005). Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens. *Science*, *309*, 1720–1722.

Meyerhöfer, W. (2006). PISA & Co als kulturindustrielle Phänomene [PISA & co. as phenomenons of culture industry]. In Th. Jahnke, & W. Meyerhöfer, (Eds.), *PISA & Co* (pp. 63–99). Hildesheim: Franzbecker.

Mullis, I. V. S., Martin, M. O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L., & Smith, T. (1997) *Mathematics achievement in the primary school years*. Chestnut Hill: TIMSS Study Center.

Mullis, I. V. S., Martin, M. O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L., & Smith, T. (1998). *Mathematics and science achievement in the final year of secondary school*. Chestnut Hill: TIMSS Study Center.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Gregory, K. D., Garden, R. A., & O'Connor, K. M., et al. (2000). *TIMSS 1999 international mathematics report*. Chestnut Hill: TIMSS Study Center.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 international mathematics report*. Chestnut Hill: TIMSS & PIRLS International Study Center.

Mullis, I., Martin, M. O., Gonzales, E. J., & Kennedy, A. M. (2003). *PIRLS 2001 international report*. Chestnut Hill: IEA.

Murray, T. S., Kirsch, I. S., & Jenkins, L. B. (1997). *Adult literacy in OECD countries*. Washington: National Center for Education Statistics.

NCES (National Center for Education Statistics). (1992). *Results of the international assessment of educational progress 1991*. Canada. Retrieved December 12, 2006 from http://nces.ed.gov/programs/digest/d96/D96T394.asp

Neuwirth, E. , Ponocny, I. , & Grossmann, W. (Eds.). (2006). *PISA 2000 und PISA 2003 [PISA 2000 and 2003]*. Wien: Leykam.

Nyborg, H. (2003). The sociology of psychometric and bio-behavioural sciences: A case study of destructive social reductionism and collective fraud in 20th century academia. In H. Nyborg (Ed.), *The scientific study of general intelligence* (pp. 441–502). Oxford: Pergamon.

OECD (2000). *Literacy in the information age. Final report of the international adult literacy survey.* Paris: OECD.

OECD (2001). *Knowledge and skills for life. First results from PISA 2000*. Paris: OECD.

OECD (2003a). *Literacy skills for the world of tomorrow. Further results from PISA 2000*. Paris: OECD.

OECD (2003b). *PISA 2003 assessment framework. Mathematics, reading, science and problem solving. Knowledge and skills*. Paris: OECD.

OECD (2004a). *Learning for tomorrow's world. First results from PISA 2003*. Paris: OECD.

OECD (2004b). *Problem solving for tomorrow's world*. Paris: OECD.

Oesterdiekhoff, G. W. (1992). *Traditionales Denken und Modernisierung, Jean Piaget und die Theorie der sozialen Evolution [Traditional thinking and modernization, Jean Piaget and the theory of social evolution]*. Opladen: Westdeutscher Verlag.

Oesterdiekhoff, G. W. (2000). *Zivilisation und Strukturgenese. Norbert Elias und Jean Piaget im Vergleich [Civilisation and genesis of structure. Norbert Elias and Jean Piaget compared].* Frankfurt: Suhrkamp.

Piaget, J. (1932). *The moral judgment of the child*. London: Kegan.

Prais, S. J. (2003). Cautions on OECD's recent educational survey (PISA). *Oxford Review of Education*, *29*, 139–163.

Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., & Neubrand, M., et al. (Eds.). (2004). *PISA 2003*. Münster: Waxmann.

Raykov, T. (2005). Analysis of longitudinal studies with missing data using covariance structure modeling with full-information maximum likelihood. *Structural Equation Modeling*, *12*, 493–505.

Rindermann, H. (2006a). Was messen internationale Schulleistungsstudien? [What do international student assessment studies measure? School performance, student abilities, cognitive abilities, knowledge or general intelligence?] *Psychologische Rundschau*, *57*, 69–86.

Rindermann, H. (2006b). Formen wissenschaftlicher Auseinandersetzung [forms of scientific disputes]. *Erwägen-Wissen-Ethik*, *17*, 251–263 u. 313–321.

Rindermann, H. (2007). Die Bedeutung der mittleren Klassenfähigkeit für das Unterrichtsgeschehen und die Entwicklung individueller Fähigkeiten [the relevance of class ability for teaching and development of individual competences]. *Unterrichtswissenschaft*, *35*, 68–89.



DOI: 10.1002/per

Rindermann, H. (in press). Relevance of education and intelligence at the national level for the economic welfare of people. *Intelligence*. DOI: 10.1016/j.intell.2007.02.002

Rindermann, H., & Neubauer, A. (2004). Processing speed, intelligence, creativity and school performance: Testing of causal hypotheses using structural equation models. *Intelligence*, *32*, 573–589.

Rushton, J. P. H. (1997/1995). *Race, evolution, and behaviour. A life history perspective*. New Brunswick: Transaction.

Rushton, J. Ph., & Jensen, A. R. (2005). Thirty years of research on Black-White differences in cognitive ability. *Psychology, Public Policy & the Law*, *11*, 235–294.

Rushton, J. Ph., Vernon, Ph. A., & Bons, T. A. (in press). No evidence that the polymorphisms of brain regulator genes Microcephalin and ASPM are associated with general mental ability, head circumference or altruism. *Biology Letters*. DOI: 10.1098/rsbl.2006.0586

Schmidt, F. L., & Hunter, J. E. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology*, *86*, 162–173.

Segerstråle, U. (2000). *Defenders of the truth. The sociobiology debate*. New York: Oxford University Press.

Shiraev, E. B., & Levy, D. A. (2004). *Cross-cultural psychology. Critical thinking and contemporary applications*. Boston: Pearson.

Simonton, D. K. (2006). Presidential IQ, openness, intellectual brilliance, and leadership: Estimates and correlations for 42 US chief executives. *Political Psychology*, *27*, 511–526.

Spearman, Ch. (1904). ''General intelligence'', objectively determined and measured. *American Journal of Psychology*, *15*, 201–292.

Stelzl, I., Merz, F., Remer, H., & Ehlers, Th. (1995). The effect of schooling on the development of fluid and cristallized intelligence: A quasi-experimental study. *Intelligence*, *21*, 279–296.

Strohschneider, S., & Güss, D. (1998). Planing and problem solving. Differences between Brazilian and German students. *Journal of Cross-Cultural Psychology*, *29*, 695–716.

Strohschneider, S., & Güss, D. (1999) The fate of the Moros: A cross-cultural exploration of strategies in complex and dynamic decision making. *International Journal of Psychology*, *34*, 235–252.

Templer, D. I., & Arikawa, H. (2006). Temperature, skin color, per capita income, and IQ: An international perspective. *Intelligence*, *34*, 121–128.

Tiedemann, J., & Billmann-Mahecha, E. (2004). Kontextfaktoren der Schulleistung im Grundschulalter. *[Contextual factors of primary school performance] Zeitschrift für Pädagogische Psychologie*, *18*, 113–124.

Vanhanen, T. (1997). *Prospects of democracy*. London: Routledge.

Vanhanen, T. (2005). *Measures of democracy 1810–2004*. Finnish Social Science Data Archive (FSD). Retrieved January 17, 2006 from www.fsd.uta.fi/english/data/catalogue/FSD1289

Wainwright, M., Wright, M. J., Geffen, G. M., Geffen, L. B., Luciano, M., & Martin, N. G. (2005). The genetic basis of academic achievement on the Queensland Core Skills Test and its shared genetic variance with IQ. *Behaviour Genetics*, *35*, 133–145.

Weede, E., & Kämpf, S. (2002). The impact of intelligence and institutional improvements on economic growth. *Kyklos*, *55*, 361–380.

Weiss, V. (2000). *Die IQ-Falle. Intelligenz, Sozialstruktur und Politik [The IQ-trap. Intelligence, social structure, and politics]*. Graz: Leopold Stocker.

Winship, Ch., & Korenman, S. (1997). Does staying in school make you smarter? The effect of education on IQ in The Bell Curve. In B. Devlin, S. E. Fienberg, D. P. Resnick, & K. Roeder (Eds.), *Intelligence, genes and success* (pp. 215–234). New York: Springer.

Woods, R. P., Freimer, N. B., De Young, J. A., Fears, S. C., Sicotte, N. L., & Service, S. K., et al. (2006). Normal variants of Microcephalin and ASPM do not account for brain size variability. *Human Molecular Genetics*, *15*, 2025–2029.

Woschek, R. (2005). *TIMSS 2 elaboriert. Eine didaktische Analyse von Schülerarbeiten im Ländervergleich Schweiz/Deutschland [TIMSS 2 elaborated. A didactic analysis of test results across Switzerland-Germany]*. Duisburg: Dissertation at University Duisburg.

Wuttke, J. (2006). Fehler, Verzerrungen, Unsicherheiten in der PISA-Auswertung [Mistakes, biases, uncertainties in PISA-interpretation]. In Th. Jahnke, W. Meyerhöfer (Eds.), *PISA & Co* (pp. 101–154). Hildesheim: Franzbecker.

Za'Rour, G. I. (1971). The conservation of number and liquid by Lebanese school children in Beirut. *Journal of Cross-Cultural Psychology*, *2*, 165–172.

 
DOI: 10.1002/per

## APPENDIX: COGNITIVE ABILITY VALUES FOR COUNTRIES

| | Unadjusted | | | | | Adjusted | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Country | TIMSS | Grade (TIMSS + PIRLS) | Age (PISA, 00-03) | Student assessment sum | All cognitive ability sum | Grade (TIMSS + PIRLS) | Age (PISA, 00-03) | Student assessment sum | All cognitive ability sum[a] | Data[b] |
| Afghanistan | | | | | 84 | | | | 78 | 0 |
| Albania | | | 386 | 449 | 87 | | 284 | 397 | 83 | 2 |
| Algeria | | | | | 83 | | | | 77 | 0 |
| Andorra | | | | | 98 | | | | 93 | 0 |
| Angola | | | | | 68 | | | | 60 | 0 |
| Antigua and Barbuda | | | | | 70 | | | | 62 | 0 |
| Argentina | | 433 | 416 | 451 | 88 | 426 | 374 | 442 | 89 | 3 |
| Armenia | 493 | 493 | | 492 | 93 | 472 | | 474 | 92 | 2 |
| Australia | 548 | 546 | 532 | 550 | 100 | 555 | 520 | 553 | 101 | 3 |
| Austria | 562 | 558 | 511 | 549 | 100 | 552 | 502 | 546 | 101 | 3 |
| Azerbaijan | | | | | 87 | | | | 81 | 0 |
| Bahamas | | | | | 84 | | | | 78 | 0 |
| Bahrain | 448 | 449 | | 447 | 85 | 459 | | 461 | 87 | 2 |
| Bangladesh | | | | | 82 | | | | 75 | 0 |
| Barbados | | | | | 80 | | | | 79 | 1 |
| Belarus | | | | | 97 | | | | 92 | 0 |
| Belgium | 534 | 538 | 517 | 536 | 99 | 539 | 515 | 541 | 100 | 3 |
| Belize | | 341 | | 337 | 75 | 339 | | 344 | 76 | 2 |
| Benin (Dahomey) | | | | | 70 | | | | 62 | 0 |
| Bermuda | | | | | 90 | | | | 90 | 1 |
| Bhutan | | | | | 80 | | | | 73 | 0 |
| Bolivia | | | | | 87 | | | | 87 | 1 |
| Bosnia | | | | | 90 | | | | 84 | 0 |
| Botswana | 396 | 398 | | 391 | 75 | 367 | | 385 | 75 | 2 |
| Brazil | | | 389 | 451 | 87 | | 312 | 416 | 84 | 3 |
| Brunei | | | | | 91 | | | | 86 | 0 |



DOI: 10.1002/per

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Bulgaria | 533 | 539 | 448 | 517 | 95 | 527 | 443 | 514 | 96 | 3 |
| Burkina Faso | | | | | 68 | | | | 60 | 0 |
| Burma (Myanmar) | | | | | 87 | | | | 81 | 0 |
| Burundi | | | | | 69 | | | | 61 | 0 |
| Cambodia (Kampuchea) | | | | | 91 | | | | 86 | 0 |
| Cameroon | | | | | 64 | | | | 61 | 1 |
| Canada | 546 | 549 | 533 | 547 | 100 | 564 | 533 | 560 | 102 | 3 |
| Cape Verde | | | | | 76 | | | | 69 | 0 |
| Central African Republic | | | | | 64 | | | | 61 | 1 |
| Chad | | | | | 68 | | | | 60 | 0 |
| Chile | 431 | 434 | 418 | 452 | 88 | 436 | 397 | 455 | 89 | 3 |
| China | | | | | 105 | | | | 107 | 1 |
| Colombia | 391 | 433 | | 431 | 84 | 377 | | 380 | 80 | 3 |
| Comoros | | | | | 77 | | | | 70 | 0 |
| Congo (Zaire) | | | | | 65 | | | | 62 | 1 |
| Congo (Brazzaville) | | | | | 64 | | | | 61 | 1 |
| Cook-Islands | | | | | 89 | | | | 89 | 1 |
| Costa Rica | | | | | 89 | | | | 83 | 0 |
| Cote d'Ivoire (Ivory Coast) | | | | | 69 | | | | 61 | 0 |
| Croatia | | | | | 90 | | | | 90 | 1 |
| Cuba | | | | | 85 | | | | 84 | 1 |
| Cyprus | 491 | 499 | | 503 | 93 | 522 | | 523 | 96 | 2 |
| Czech Republic | 562 | 554 | 511 | 547 | 99 | 551 | 508 | 547 | 100 | 3 |
| Denmark | 492 | 506 | 504 | 520 | 97 | 532 | 501 | 536 | 99 | 3 |
| Djibouti | | | | | 68 | | | | 60 | 0 |
| Dominica | | | | | 67 | | | | 64 | 1 |
| Dominican Republic | | | | | 82 | | | | 81 | 1 |
| Ecuador | | | | | 88 | | | | 88 | 1 |
| Egypt | 442 | 443 | | 442 | 84 | 441 | | 443 | 84 | 3 |
| El Salvador | | | | | 80 | | | | 73 | 0 |
| Equatorial Guinea | | | | | 59 | | | | 55 | 1 |
| Eritrea | | | | | 68 | | | | 60 | 0 |
| Estonia | 564 | 563 | | 564 | 101 | 531 | | 531 | 99 | 3 |

*Continues*




DOI: 10.1002/per

APPENDIX (Continued)

| Country | Unadjusted | | | | | Adjusted | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | TIMSS | Grade (TIMSS + PIRLS) | Age (PISA, 00-03) | Student assessment sum | All cognitive ability sum | Grade (TIMSS + PIRLS) | Age (PISA, 00-03) | Student assessment sum | All cognitive ability sum[a] | Data[b] |
| Ethiopia | | | | | 64 | | | | 61 | 1 |
| Fiji | | | | | 85 | | | | 84 | 1 |
| Finland | 548 | 547 | 546 | 559 | 101 | 570 | 547 | 570 | 103 | 3 |
| France | 523 | 533 | 513 | 540 | 99 | 531 | 515 | 540 | 100 | 3 |
| Gabon | | | | | 64 | | | | 55 | 0 |
| Gambia | | | | | 66 | | | | 57 | 0 |
| Georgia | | | | | 94 | | | | 89 | 0 |
| Germany | 525 | 541 | 501 | 535 | 99 | 519 | 498 | 532 | 99 | 3 |
| Ghana | 301 | 304 | | 300 | 67 | 256 | | 263 | 64 | 3 |
| Greece | 494 | 522 | 470 | 517 | 95 | 546 | 465 | 534 | 97 | 3 |
| Grenada | | | | | 71 | | | | 63 | 0 |
| Guatemala | | | | | 79 | | | | 78 | 1 |
| Guinea | | | | | 67 | | | | 64 | 1 |
| Guinea-Bissau | | | | | 67 | | | | 59 | 0 |
| Guyana | | | | | 87 | | | | 81 | 0 |
| Haiti | | | | | 67 | | | | 59 | 0 |
| Honduras | | | | | 81 | | | | 80 | 1 |
| Hong Kong | 574 | 563 | 542 | 558 | 104 | 569 | 538 | 561 | 106 | 3 |
| Hungary | 555 | 554 | 497 | 540 | 99 | 549 | 491 | 544 | 100 | 3 |
| Iceland | 491 | 515 | 509 | 529 | 99 | 544 | 509 | 545 | 101 | 3 |
| India | | | | | 82 | | | | 81 | 1 |
| Indonesia | 445 | 447 | 390 | 446 | 86 | 439 | 338 | 440 | 86 | 3 |
| Iran | 451 | 453 | | 452 | 86 | 439 | | 442 | 85 | 3 |
| Iraq | | | | | 87 | | | | 87 | 1 |
| Ireland | 544 | 545 | 515 | 540 | 97 | 539 | 511 | 542 | 98 | 3 |
| Israel | 511 | 513 | 451 | 505 | 94 | 523 | 425 | 506 | 96 | 3 |
| Italy | 517 | 529 | 482 | 526 | 99 | 550 | 481 | 540 | 101 | 3 |
| Jamaica | | | | | 71 | | | | 69 | 1 |



DOI: 10.1002/per

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Japan | 585 | 579 | 540 | 570 | 104 | 580 | 539 | 572 | 105 | 3 |
| Jordan | 471 | 472 | | 471 | 87 | 488 | | 489 | 90 | 3 |
| Kazakhstan | | | | | 94 | | | | 89 | 0 |
| Kenya | | | | | 72 | | | | 70 | 1 |
| Kiribati | | | | | 85 | | | | 79 | 0 |
| Korea (North) | | | | | 106 | | | | 102 | 0 |
| Korea (South) | 595 | 589 | 544 | 576 | 105 | 588 | 531 | 573 | 106 | 3 |
| Kuwait | 400 | 424 | | 422 | 83 | 406 | | 409 | 83 | 3 |
| Kyrgyzstan | | | | | 90 | | | | 84 | 0 |
| Laos | | | | | 89 | | | | 89 | 1 |
| Latvia | 525 | 536 | 481 | 527 | 97 | 520 | 476 | 522 | 98 | 2 |
| Lebanon | 442 | 443 | | 441 | 84 | 432 | | 435 | 84 | 3 |
| Lesotho | | | | | 67 | | | | 59 | 0 |
| Liberia | | | | | 67 | | | | 59 | 0 |
| Libya | | | | | 83 | | | | 77 | 0 |
| Liechtenstein | | | 515 | 541 | 99 | | 485 | 528 | 99 | 2 |
| Lithuania | 509 | 524 | | 525 | 95 | 507 | | 508 | 94 | 3 |
| Luxembourg | | | 474 | 512 | 97 | | 475 | 522 | 99 | 2 |
| Macau | | | 524 | 547 | 100 | | 492 | 533 | 99 | 2 |
| Macedonia | 473 | 468 | 401 | 463 | 89 | 452 | 332 | 441 | 88 | 2 |
| Madagascar | | | | | 82 | | | | 81 | 1 |
| Malawi | | | | | 69 | | | | 61 | 0 |
| Malaysia | 530 | 530 | | 530 | 96 | 531 | | 531 | 97 | 3 |
| Maldives | | | | | 81 | | | | 74 | 0 |
| Mali | | | | | 69 | | | | 61 | 0 |
| Malta | | | | | 97 | | | | 98 | 1 |
| Marianas | | | | | 81 | | | | 80 | 1 |
| Marshall Islands | | | | | 84 | | | | 83 | 1 |
| Mauritania | | | | | 76 | | | | 69 | 0 |
| Mauritius | | | | | 89 | | | | 89 | 1 |
| Mexico | | | 415 | 470 | 89 | | 330 | 427 | 85 | 3 |
| Micronesia | | | | | 84 | | | | 78 | 0 |
| Moldova | 500 | 501 | | 501 | 94 | 486 | | 487 | 94 | 2 |

*Continues*




DOI: 10.1002/per

APPENDIX (Continued)

| Country | Unadjusted | | | | | Adjusted | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | TIMSS | Grade (TIMSS + PIRLS) | Age (PISA, 00-03) | Student assessment sum | All cognitive ability sum | Grade (TIMSS + PIRLS) | Age (PISA, 00-03) | Student assessment sum | All cognitive ability sum[a] | Data[b] |
| Mongolia | | | | | 101 | | | | 97 | 0 |
| Morocco | 386 | 382 | | 379 | 79 | 351 | | 355 | 77 | 3 |
| Mozambique | | | | | 64 | | | | 61 | 1 |
| Namibia | | | | | 70 | | | | 62 | 0 |
| Nepal | | | | | 78 | | | | 77 | 1 |
| Netherlands | 561 | 560 | 527 | 548 | 100 | 565 | 528 | 557 | 102 | 3 |
| New Caledonia | | | | | 85 | | | | 84 | 1 |
| New Zealand | 525 | 530 | 532 | 543 | 99 | 544 | 525 | 549 | 101 | 3 |
| Nicaragua | | | | | 81 | | | | 74 | 0 |
| Niger | | | | | 69 | | | | 61 | 0 |
| Nigeria | | | | 439 | 79 | | | 388 | 75 | 3 |
| Norway | 510 | 513 | 503 | 525 | 98 | 530 | 502 | 536 | 100 | 3 |
| Oman | | | | | 83 | | | | 77 | 0 |
| Pakistan | | | | | 84 | | | | 83 | 1 |
| Panama | | | | | 84 | | | | 78 | 0 |
| Papua New Guinea | | | | | 83 | | | | 82 | 1 |
| Paraguay | | | | | 84 | | | | 83 | 1 |
| Peru | | | 339 | 416 | 82 | | 279 | 394 | 81 | 3 |
| Philippines | 394 | 398 | | 418 | 83 | 392 | | 425 | 85 | 3 |
| Poland | | | 492 | 525 | 98 | | 487 | 530 | 99 | 3 |
| Portugal | 472 | 492 | 474 | 505 | 94 | 476 | 463 | 497 | 95 | 3 |
| Puerto Rico | | | | | 84 | | | | 83 | 1 |
| Qatar | | | | | 78 | | | | 77 | 1 |
| Romania | 493 | 505 | | 505 | 94 | 479 | | 481 | 93 | 3 |
| Russia | 545 | 543 | 476 | 529 | 97 | 550 | 472 | 535 | 99 | 3 |
| Rwanda | | | | | 70 | | | | 62 | 0 |
| Saint Kitts and Nevis | | | | | 67 | | | | 59 | 0 |
| Saint Lucia | | | | | 62 | | | | 59 | 1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Saint Vincent, Grenadines | | | | | 71 | | | | 69 | 1 |
| Samoa (Western Samoa) | | | | | 88 | | | | 88 | 1 |
| Sao Tome and Principe | | | | | 67 | | | | 59 | 0 |
| Saudi Arabia | 396 | 398 | | 395 | 80 | 407 | | 410 | 82 | 2 |
| Senegal | | | | | 66 | | | | 57 | 0 |
| Seychelles | | | | | 86 | | | | 80 | 0 |
| Sierra Leone | | | | | 64 | | | | 61 | 1 |
| Singapore | 607 | 584 | | 570 | 105 | 588 | | 578 | 107 | 3 |
| Slovakia | 548 | 542 | 494 | 535 | 98 | 546 | 488 | 538 | 99 | 3 |
| Slovenia | 544 | 535 | | 535 | 98 | 533 | | 537 | 99 | 3 |
| Solomon Islands | | | | | 84 | | | | 78 | 0 |
| Somalia | | | | | 68 | | | | 60 | 0 |
| South Africa | 304 | 328 | | 324 | 69 | 281 | | 287 | 66 | 3 |
| Spain | 505 | 516 | 492 | 520 | 97 | 511 | 483 | 523 | 98 | 3 |
| Sri Lanka | | | | | 79 | | | | 78 | 1 |
| Sudan | | | | | 71 | | | | 69 | 1 |
| Suriname | | | | | 89 | | | | 89 | 1 |
| Swaziland | | | | | 68 | | | | 60 | 0 |
| Sweden | 534 | 548 | 516 | 547 | 100 | 543 | 517 | 549 | 101 | 3 |
| Switzerland | 540 | 542 | 516 | 541 | 100 | 546 | 512 | 544 | 101 | 3 |
| Syria | 414 | 416 | | 414 | 82 | 429 | | 432 | 84 | 3 |
| Taiwan | 590 | 588 | | 589 | 106 | 596 | | 595 | 108 | 3 |
| Tajikistan | | | | | 87 | | | | 81 | 0 |
| Tanzania | | | | | 72 | | | | 70 | 1 |
| Thailand | 502 | 508 | 439 | 498 | 92 | 497 | 404 | 480 | 91 | 3 |
| Timor-Leste | | | | | 87 | | | | 81 | 0 |
| Togo | | | | | 70 | | | | 62 | 0 |
| Tonga | | | | | 86 | | | | 86 | 1 |
| Trinidad and Tobago | | | | 491 | 90 | | | 518 | 93 | 2 |
| Tunisia | 441 | 443 | 379 | 443 | 84 | 428 | 376 | 444 | 85 | 2 |
| Turkey | 457 | 461 | 436 | 472 | 90 | 462 | 346 | 451 | 88 | 3 |
| Turkmenistan | | | | | 87 | | | | 81 | 0 |
| Uganda | | | | | 73 | | | | 71 | 1 |

*Continues*





DOI: 10.1002/per

APPENDIX (Continued)

| Country | Unadjusted | | | | | Adjusted | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | TIMSS | Grade (TIMSS + PIRLS) | Age (PISA, 00-03) | Student assessment sum | All cognitive ability sum | Grade (TIMSS + PIRLS) | Age (PISA, 00-03) | Student assessment sum | All cognitive ability sum[a] | Data[b] |
| Ukraine | | | | | 97 | | | | 92 | 0 |
| United Arab Emirates | | | | | 84 | | | | 78 | 0 |
| United Kingdom | 535 | 540 | 533 | 547 | 100 | 555 | 528 | 556 | 102 | 3 |
| United States | 538 | 542 | 499 | 540 | 99 | 549 | 499 | 540 | 100 | 3 |
| Uruguay | | | 436 | 484 | 93 | | 384 | 463 | 92 | 3 |
| Uzbekistan | | | | | 87 | | | | 81 | 0 |
| Vanuatu | | | | | 84 | | | | 78 | 0 |
| Venezuela | | | | 439 | 84 | | | 437 | 85 | 3 |
| Vietnam | | | | | 94 | | | | 95 | 1 |
| Yemen | 354 | 356 | | 353 | 76 | 327 | | 331 | 75 | 3 |
| Yugoslavia (Serbia) | 498 | 498 | 436 | 491 | 91 | 477 | 423 | 483 | 91 | 3 |
| Zambia | | | | | 71 | | | | 69 | 1 |
| Zimbabwe | | | | 416 | 76 | | | 384 | 73 | 3 |
| **Mean** | **494** | **495** | **479** | **491** | **84** | **491** | **460** | **486** | **82** | |
| Standard deviation | 68 | 67 | 52 | 63 | 12 | 77 | 74 | 72 | 14 | |
| *N* | 63 | 65 | 48 | 78 | 194 | 65 | 48 | 78 | 194 | |

*Note*: The *all cognitive ability sum* was standardised according to Lynn and Vanhanen (IQ = 100 for UK, 'Greenwich IQ', $M_{UK} = 100$, $SD_{UK} = 15$); for the unadjusted student assessment studies the given norms were used (usually OECD-countries of the developed world, $M = 500$, $SD = 100$); for the adjusted sum, the subtractions in the three population measures PISA00corrected, PISA03corrected and IQ-test-collection corrected are considered.

[a]Obtained for 194 countries (measured data for 128 countries, 66%, estimated for 66, 34%).

[b]3, both (student assessment and intelligence test studies, 32.5%); 2, student (only student assessment studies, 7.7%); 1, IQ (only intelligence test studies, 25.8%); 0, estimated (34%).