European Journal of Personality
Eur. J. Pers. **21**: 707–765 (2007)
Published online in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/per.648

Discussion on 'The *g*-Factor of International Cognitive Ability Comparisons: The Homogeneity of Results in PISA, TIMSS, PIRLS and IQ-Tests Across Nations' by Heiner Rindermann

OPEN PEER COMMENTARY

Geographical Distribution of Mental Abilities and Its Moral Consequences

JÜRI ALLIK

Department of Psychology, University of Tartu, Estonia
Juri.Allik@ut.ee

Abstract

Rindermann's study provides the most comprehensive evidence so far that national scores of school assessment have systematic differences and the geographical distribution of these differences almost perfectly repeat the distribution of the mean national scores of intelligence. It is argued that without comparison with the random effects of statistical aggregation it is impossible to decide whether additional factors are needed to explain the strong association between national scores of school assessment and intelligence tests. The ignorance about real differences in mental abilities may become a source of social injustice because this does not allow natural inequalities to be arranged such that they are to the greatest benefit of the least advantageous. Copyright © 2007 John Wiley & Sons, Ltd.

Rindermann (this issue) provides the most comprehensive evidence so far that national scores of school assessment have systematic differences and the geographical distribution of these differences almost perfectly repeat the distribution of the mean national scores of intelligence. By analogy with many previous controversial discoveries, it is predictable that the first most typical reaction would be denial. Indeed, many critics are not able to tolerate the idea that the mean level of intelligence could systematically vary across countries and world regions. Neither they are ready to accept that from the distribution of mental resources it is possible to predict the wealth of nations, as Lynn and Vanhanen (2002) did in their widely acclaimed book 'IQ and the wealth of nations'. This cannot be true because we have taught and believe otherwise (Volken, 2003).

The next predictable phase is acceptance of the facts but denying their interpretation. The simplest strategy is to interpret the results as measurement error. A useful strategy is to

discover few small mistakes declaring all other results equally suspicious (Kamin, 2006). The main implication of Rindermann's paper is that this strategy of denial is not effective any more. Like personality traits (Allik & McCrae, 2004), the geographical distribution of intelligence is robust and even surprisingly immune to all sorts of measurement error.

Level of analysis. Rindermann assumes that factors contributing to the high correlation between school assessment and intelligence scores at the individual level are also responsible for equally high or even larger correlations at the national level. In addition, as Rindermann notices, there may be factors exclusive for the national level alone.

Technically speaking, the aggregate level (e.g. national) correlation between two variables may not repeat exactly the individual level of correlation: it may be larger or smaller and even of opposite sign (Ostroff, 1993). Perhaps the best example is the relationship between life satisfaction and suicide. At the individual level dissatisfaction with one's life is a strong predictor of suicidal tendencies: it is disappointing and troubles people who are inclined to commit suicide (Koivumaa-Honkanen, Honkanen, Viinamäki, Heikkila, Kaprio, & Koskenvuo, 2001). Nevertheless, looking at national-level indicators the suicide rate is systematically higher in countries where people on average are more happy and satisfied with their lives. Thus, at the country level the correlation between life satisfaction and suicide is strongly positive (Inglehart, 1990). This is a good example what is usually called 'ecological fallacy' (see Inglehart & Welzel, 2005, Chap. 10). This example suggests that theoretically it is possible that student assessment and intelligence scores are strongly correlated at the individual level of analysis but loose this relationship or even reverse it at the aggregated national level of analysis. Rindermann's analyses demonstrated that it is not so: The correlation between the national scores of intelligence and school assessment is even higher than at the individual level.

Beside assumptions that some common factors operate at both levels it is necessary to take into account simple statistical consequences of aggregation. If there is a strong correlation between variables *X* and *Y* at the individual level, this correlation generally preserves and even increases when individuals are randomly assigned to artificially created groups and the correlation is computed between mean values of X and Y of these groups (see McCrae, Terracciano, & 79 Members of the Personality Profiles of Cultures Project, 2005). In other words, correlations between variables would be preserved at the aggregate level in this case, without assumption that factors instrumental at the individual level continue their operation at the level of the whole nation. A clear implication is that without comparison with the random effects of statistical aggregation it is impossible to decide whether additional factor are needed to explain the strong association between national scores of school assessment and intelligence tests.

Ethical problems. The opposition to Lynn and Vanhanen (2002, 2006) has been predominantly ethical in nature. One of the deepest and basic principles on which the Western societies are built on is the presumption that all people are born equal. There is a strong tendency to believe that they are born equal not only in terms of the same rights but also in terms of capacities and potentials. Questioning these fundamental principles is perceived as threat to the human society and its underlying principles. Therefore it seems that concerns are not so much about accuracy of scientific facts than about a fear that fundamental moral principles are violated. Because ethical principles are above scientific truth, the mere idea that a group of people is less successful than some other group, not due to social or historical circumstances but to natural propensities, is intolerable. Recent history is full of examples of injustice and even crimes against humanity done in the name

DOI: 10.1002/per

of dogma that one racial, ethnical or religious group is superior of all others. Anything that even could hint on this possibility seems morally unacceptable.

But let us assume that human cognitive abilities vary from one human group to another like many other biological and anthropometrical attributes. One moral imperative, just explicated above, tells us to ignore these differences. It is better to put oneself, as Rawls (1999) suggested, behind a veil of ignorance. Behind this veil, one knows nothing of natural abilities or anyone's position in society. Behind such a veil of ignorance all individuals are simply specified as rational, free and morally equal beings (Rawls, 1999). Although attractive and even inevitable as a basis of social contract, this moral position contains a possibility to commit another kind of unfairness by ignoring really existing ability differences. Behind the veil of ignorance it would be difficult if not possible to satisfy conditions of Rawls's second principle: All social values are to be distributed equally unless an unequal distribution of any, or all, of these values is to everyone's advantage (Rawls, 1999, p. 54). It may be so that to everyone's advantage it would be necessary to arrange natural abilities such that they are to the greatest benefit of the least advantageous. How can intellectual inequalities be arranged if we are ignorant of them and are not able to measure them? Thus, a lift of the veil of ignorance is tolerable when it is necessary to avoid an even greater injustice.

The Big *G*-Factor of National Cognitive-Ability Comparisons: Not Trivial and Not Immutable

JENS B. ASENDORPF

Department of Psychology, Humboldt University Berlin, Germany
jens.asendorpf@rz.hu-berlin.de

Abstract

Rindermann's analysis identifies a large first factor of cross-national differences, with high loadings of both IQ tests and student achievement tests. This finding is not trivial because correlations at different levels of analysis such as individuals and nations can be very different. The finding should be seriously discussed rather than downplayed because of the enormous political implications if the finding is misinterpreted as evidence for immutable national or regional differences. Copyright © 2007 John Wiley & Sons, Ltd.

An earlier German contribution by Rindermann (2006) already aroused some debate among German scholars (Baumert, Brunner, Lüdtke, & Trautwein, 2007; Prenzel, Walter, & Frey, 2007), and I expect that the present target paper that focuses only on cross-national comparisons, including new data and new analyses, will raise even more eyebrows, and invite criticism. I welcome this development because, as one reviewer of the target paper put it, some 'student assessment researchers are trying to monopolise interpretation, publication and their point of view attempting to build an impenetrable wall between assessment of academic achievements and psychometric IQ studies . . . This wall was erected not by the strength of logical arguments but mainly by fears to violate rules of political correctness'.

DOI: 10.1002/per

My commentary focuses on two main potential misinterpretations of the target paper's take home message that, at the national level, student achievement tests such as PISA and TIMSS assess, by and large, national general intelligence.

First, the message is *not trivial* because correlations between two individual difference variables such as achievement in PISA versus an intelligence test do not necessarily become stronger, due to reduction of measurement error, when means of populations rather than individual scores are correlated. That they would become stronger seems to be a widespread assumption, but it is an erroneous one. The correlations may even change sign. Consider Figure 1(A). This illustrates Rindermann's findings of a high positive correlation between PISA and general intelligence means across nations on the basis of positive (but perhaps not just as strong) correlations across individuals within nations.

Now consider the principally possible case illustrated by Figure 1(B) that highly positive correlations apply within nations along with a highly *negative* correlation across

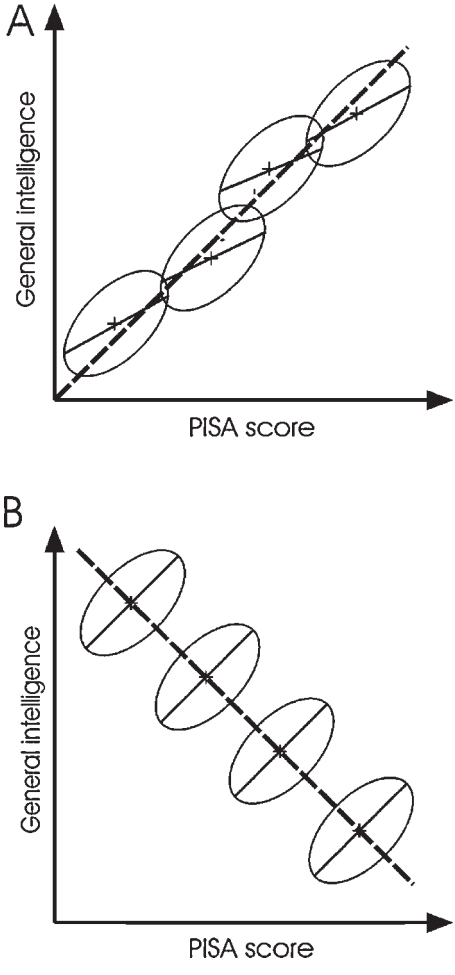

Figure 1. Principle independence of correlations within nations (regression lines within ellipses) and between nations (regression lines for national means) (adapted from Asendorpf, 2004, Figure 8.1). Panel A: Illustration of Rindermann's (this issue) finding. Panel B: Fictitious data with positive correlations within but negative correlation between nations.

DOI: 10.1002/per

nations. The principal independence of correlations within and between cultures has been recognised by some sociologists, psychologists and behavioural geneticists for a long time (e.g. Asendorpf, 2004, Chap. 8; Leung & Bond, 1989; Ostroff, 1993; Robinson, 1950; Rowe, Vazsonyi, & Flannery, 1994). A structurally identical problem at a lower level of analysis is the principal independence of intra-individual correlations such as a negative correlation between the states of being angry and happy across situations, and inter-individual correlations such as a slightly positive inter-individual correlation between the traits of angriness and happiness across individuals (Epstein, 1983; Molenaar, 2004).

The principle independence of correlations within and across nations is due to the fact that the mechanisms that make individuals different from one another are by no means identical with the mechanisms that make nations different from one another, and that the differences between the mechanisms affect the similarity of the within- and between-nation correlations. A thought example may help to better understand this point. Assume that the educational system is identical across nations, and only 50% of the population attend school at all within each nation. In this case, schooling is a variable that influences test achievement within nations but not between nations. Because the educational system is identical across nations, the cross-national differences in intelligence and PISA are expected to be lower than Rindermann's (this issue) results tell us, and because only 50% of the population attend school, the within-nation differences in intelligence and PISA are expected to be higher. Accordingly, the correlation between intelligence and PISA is expected to be lower across nations than within nations. Consequently, finding a large first factor of the mean cognitive ability of nations is by no means trivial. Instead, it is an important empirical finding that cannot be deduced from studies within nations. In order to avoid confusion I suggest here to distinguish a *Big* G-*Factor* of cross-national comparisons in cognitive ability from Spearman's (1904) *g-factor* of cross-individual comparisons in cognitive ability.

My suspicion is that educational researchers currently try to downplay this finding, anxiously avoiding to relate their results to intelligence test achievement, because they fear that it would be considered politically incorrect, undermining funding of large-scale studies such as PISA or TIMSS. This may be true but trying to avoid the problem instead of solving it may be a bad advice, at least in the long run. The problem is not the fact of high cross-national correlations itself but the risk of misinterpretation of this result. A similar discussion has plagued research on racial differences in IQ in the USA for quite some time. I believe that the main arguments from that discussion can be applied here as well: Differences in intelligence and student achievement across nations are small relative to the differences within nations; they are susceptible to historical change; they are not immutable but susceptible to educational and political intervention to some degree; results on the considerable influence of genetic differences on intelligence (and probably also PISA and TIMSS test scores) within nations cannot be generalised to the influence of genetic differences between nations on differences in national cognitive ability (simply replace PISA by measured genetic differences, and Figure 1 applies here as well). As scientists, we are obliged to use our competence to make these and related arguments as clear as possible to the public. If we succeed, national differences in intelligence will be better understood, and not be perceived as politically incorrect but rather as a call for political and educational interventions designed to increase the national intelligence levels of the lower scoring nations.

Last but not least, I like to draw attention to regional differences in intelligence and student achievement within nations, a point not discussed by Rindermann (this issue) (but see Rindermann, 2006). Such differences are usually small, but significant in large-scale

DOI: 10.1002/per

studies, and the results by Rindermann (this issue) suggest, although not prove, that they apply also to regional differences in some cases (mechanisms creating differences between regions within a nation can be different from mechanisms creating differences between nations). Are Southern Germans (slightly) more intelligent than Northern Germans, are Southern Italians (slightly) less intelligent than Northern Italians? Probably yes, and avoiding this discussion does not help to solve this problem. More affluent regions require and attract more intelligent labour force, and because moving within a nation is easier than to emigrate, such personality—environment correlations are therefore even more likely to build up—and to become a target for intervention.

The Different Levels of the *g*-Factor

ROEL J. BOSKER

Institute for Educational Research, University of Groningen, The Netherlands
r.j.bosker@rug.nl

Abstract

In search for a country-level g-factor, aggregated data were used by Rindermann. This, however, can cause some problems, well-known in the literature on multi-level modelling. Conceptual clarity of the country-level g-factor is lacking, the 'blown up' coefficients are well-known in studies on using aggregated data, and are not that remarkable, and the quest for causes and effects of the country-level g-factor is only meaningful if we understand what this factor stands for, and if we are able to analyse data in which the multi-level structure is preserved. Copyright © 2007 John Wiley & Sons, Ltd.

Using aggregated data can be prone with errors. In the literature on multi-level modelling (e.g. Goldstein, 2003; Kreft, 1987; Kreft & de Leeuw, 1998; Longford, 1993; Raudenbush & Bryk, 2005; Snijders & Bosker, 1999) some errors are discussed that are related to Rindermann's study. Do these errors occur in this study? Moreover I will make some more general methodological remarks that do not relate to using aggregate level analyses as such.

Shift of meaning. Using aggregates of variables, one should be aware of the fact that such aggregates generally have a different conceptual meaning than the individual-level variables. Class level IQ, for instance, might conceptually be interpreted as the richness of the learning environment. This of course is something quite different from the individual level meaning of IQ. In Rindermann's study, one cannot help but ask the question what this country-level *g*-factor actually means. Is it the ability of a society to deal with complex societal tasks and create prosperity? Or is it just the sum of the abilities of individuals, and is prosperity of a society just the result of individual efforts? And if it is the ability of a society to solve complex societal tasks, should not that ability be assessed by looking at the cognitive abilities of the elite only? This shift of meaning also causes problems in interpreting country-level relationships, such as between democracy, wealth, gross national product and level of education of the population on one hand and this country-level *g*-factor

DOI: 10.1002/per

on the other. What is one explaining here? Finding some empirical support for convergent validity of the content-specific scales at the country level does not solve this conceptual problem. Rindermann's study would clearly had more value, had these conceptual issues been addressed more thoroughly.

Aggregation bias and the ecological fallacy. Already in 1950 Robinson wrote about the problem that we cannot make straightforward inferences from aggregated data about individual behaviour. The most well-known example is that in areas where many ethnic minorities are living, people tend to vote for right wing political parties. One cannot infer from this observed relationship that ethnic minorities vote more right wing. If one still does, the ecological fallacy occurs. The reverse is also true: we cannot infer from individual-level relationships about the strength nor the size of group level relationships. Robinson (1950) also demonstrated that if we aggregate data, the correlation between group means will generally exceed the total correlation, ignoring the grouping, between the individual-level variables. This is also known as a 'blown up' correlation, and using such an estimate to make inferences about individual-level relationships is hindered by what is called aggregation bias. This is even truer if group means are measured with almost perfect reliability, which is the case in the assessment studies used by Rindermann. Reliability of aggregated variables is namely a function of the number of individual-level units, in this case mostly 1000 students or more per country.

That there is a strong *g*-factor at the country level is in itself therefore not a surprising result, but of course it can only occur if there is a cause for a country-level factor structure. I will return to this issue later on. In addition to a conceptual clarity with respect to the country-level *g*-factor, Rindermann's study suffers even more from a mix of questions, statements and explanations of individual and country-level relationships. The paper starts with a summary of findings that both relate to a person's environment, his cognitive ability and performance as well as to a country's mean IQ and economic wealth. Then the author states that 'the conceptual and developmental associations between intelligence and student achievement have been neglected' and continues that 'at the level of international and cross-cultural comparisons, the relationships between different measures of cognitive ability . . . have been largely disregarded'. The questions put forward are generally stated at the country level, but the third question is looking for a cause of a general *g*-factor lying in the cognitive demands of the tasks, and that of course relates to a question at the level of the individual students where coincidentally the structure of the *g*-factor turns out to be weaker than at the country level. In the Explanations section the author returns to the findings that relate to this question, and also puts forward a genetic explanation. But these student-level explanations do not help us in explaining the more prominent existence of a country-level *g*-factor as such.

The impossibility of estimating cross-level effects. By aggregating data the variability within the countries is lost. This then makes it impossible to look for across level relations, such as that a country's mean IQ might affect student's science literacy. Or that the relation between a student's family background and his or her cognitive ability may be moderated by a country's *g*-factor. Rindermann does not phrase questions about the existence of such relationships, yet he puts forward as a possible explanation of his findings that the intelligence of others creates intelligence. At least some of the data sets, for instance the PISA and IEA assessment studies, allow for the testing of such an explanation, but then only if the data are analysed in a simultaneous two-level model in which the original data structure is preserved. Such a data structure is also needed if future studies address Rindermann's assertion 'that individual development of cognitive abilities depends on macro-social conditions of

DOI: 10.1002/per

societies'. A general observation in the social sciences is that the environment of an individual looses strength of impact the more distal the environment is. Or stated reversely: A child's cognitive development is by far more dependent on the family, peer, neighbourhood and school context, than on regional and country-level factors.

Model misspecification. In 'Context and consex: a cautionary tale', Hauser (1970) has pointed to the risks of finding interesting macro-level phenomena and making all sorts of inferences based on this finding, whereas in actual fact the phenomenon only occurs because of model misspecification at the micro-level. This objection, however, does not pertain to the study at hand. It would, however, be worthwhile to acknowledge this in future studies in which the effects of individual-level variables on the student-level *g*-factor, as mentioned by Rindermann, might be partialled out in looking for the 'residual country-level *g*-factor' and its country-level correlates.

Missing data. A more general remark is that Rindermann in some cases is using assessment data from 48 countries only, whereas country-level IQ measures are available for 194 countries. Using full information maximum likelihood to deal with pairwise missing data at the country level does not solve the problem of non-random causes of the missing data process.

Not Every g is *g*

MARTIN BRUNNER and ROMAIN MARTIN

Research Unit for Educational Measurement and Applied Cognitive Science, University of Luxembourg, Luxembourg
martin.brunner@emacs.lu; romain.martin@emacs.lu

Abstract

The target paper identifies a common factor underlying measures of intelligence and student achievement on the cross-national level. Given the level of analysis applied, however, this factor cannot be interpreted as general cognitive ability (g). Rather, it is an indicator of a nation's prosperity. g operates at the individual level and not at the cross-national level. Copyright © 2007 John Wiley & Sons, Ltd.

It is scientifically fruitful to investigate the relationship between general cognitive ability (*g*, as measured in intelligence tests) and students' subject-specific or domain-specific abilities (e.g. as measured in the PISA tests, see Messick, 1984). Unfortunately, however, the analyses provided by Rindermann (this issue) do not permit this relationship to be quantified, or any conclusions to be drawn about whether or not *g* (which Rindermann refers to as 'ability to think' or 'intelligence') is identical to student achievement, on either a macro-social level or an individual level.

The target paper draws on aggregated data from 173 nations; however, for many of these nations, data are not available for some or all of the measures applied. Moreover, although aware of some of its limitations (Hunt & Sternberg, 2006; Mackintosh, 2007), Rindermann uses the data base by Lynn and Vanhanen (2002, 2006), which uses different intelligence measures with different metrics across nations, thus failing to fulfil minimum requirements

DOI: 10.1002/per

for measurement invariance (Meredith, 1993). In sum, the data used in the target paper suffer serious methodological limitations.

Even if we assume Rindermann's data base to be acceptable, is it surprising that a common factor was found for measures of intelligence and student achievement on the cross-national level? We do not think so. The socio-economic status of nations is known to differ substantially. Recent cross-national investigations have shown that measures of cognitive performance (e.g. intelligence tests or PISA tests) are strongly related to indicators of national prosperity, such as gross domestic product *per capita* (Hunt & Wittmann, in press). This relationship may be best explained through reciprocal causation. On one hand, the cognitive performance of a nation's population contributes to that nation's wealth. On the other hand, wealthier nations provide their citizens not only with better basic conditions for positive cognitive development (e.g. better nutrition), but also with better learning opportunities, such as more cognitively stimulating environments in families, a better cultural infrastructure and, last but not least, better schools. There is now a wealth of empirical evidence to show that better schooling enhances cognitive processes related to both specific content areas and general cognitive ability (e.g. Ceci, 1991). Consequently, given that *nations* differ considerably in the school-related (and school-unrelated) learning opportunities they provide, substantial *mean* differences in achievement measures and measures of *g* can be expected *across nations*.

It follows that a strong common factor can be extracted for such measures when factor analysis is performed at the cross-national level. The common factor found by Rindermann reflects the strong relationship that operates at the cross-national level between socio-economic and socio-cultural specificities (best summarised by indicators of national prosperity), on one hand, and various cognitive measures, on the other. In sum, disregarding for the moment the major methodological drawbacks of the data, the analyses presented in the target paper may identify a common factor underlying measures of intelligence and student achievement at the cross-national level. Crucially, however, this common factor can only be interpreted at the same level as the level of analysis adopted, namely nations. From this perspective, the factor may be interpreted as another indicator of a nation's prosperity or socio-cultural wealth. Hence, we strongly disagree with Rindermann's conclusion—as suggested primarily in the Introduction and Discussion sections of the paper—that this factor reflects 'the ability to think' or 'knowledge'.

Cognitive abilities are commonly considered to operate at the *individual level* (Baumert et al., 2007; Carroll, 1993; Neisser et al., 1996). The aggregation of individual students' scores on *g* or measures of student achievement on the national level does *not* imply that *g* or student achievement are constructs that operate at the cross-national level. The nation mean is a summary statistic that reflects the average level of domain-specific student achievement or the level of *g* achieved by the individual students in a particular nation. Moreover, it has been established for over 100 years now that individuals' scores on measures of cognitive abilities are positively correlated across domains (Gustafsson & Undheim, 1996). Contemporary psychometric theories (Neisser et al., 1996) on the structure of cognitive abilities attribute this 'positive manifold' to latent variables representing both *g* and abilities that are specific to domains or cognitive operations (Carroll, 1993). And yes, when structural theories of cognitive abilities are applied, *individual* students' domain-specific abilities and *g* explain substantial amounts of the variance in PISA measures (Brunner, 2005, 2006).

This has serious implications for Rindermann's conclusions. Interpreting the data used in the target paper from a two-level modelling perspective implies that the measures

DOI: 10.1002/per

applied show a positive manifold within and between nations (cf. Härnqvist, Gustafsson, Muthen, & Nelson, 1994). Consequently, it is possible to identify two common factors—one within each nation and one across nations—that are mutually independent. According to current thinking on cognitive abilities, the individual-level factor represents *g*. Mainstream intelligence research defines individuals' *g* as 'a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, [and] comprehend complex ideas'. (Gottfredson, 1997b, p. 13). There is no systematic relationship between individuals' *g* and the common factor across nations. It thus remains to ask why the construct description for the cross-national factor given by Rindermann overlaps to such an extent with the mainstream interpretation of individuals' *g*. Do nations cognise? We and others (e.g. James, Joyce, & Slocum, 1988) do not think so. Correlations on the cross-national level do not allow substantive inferences to be drawn on inter-individual differences in individuals' cognitive processes (Ostroff, 1993; Robinson, 1950). Hence, the common factor identified on the cross-national level is *not* aligned to individuals' general cognitive ability. In other words, not every g is *g*.

Little *g*: Prospects and Constraints

STEPHEN J. CECI and WENDY M. WILLIAMS

Department of Human Development, Cornell University, USA
sjc9@cornell.edu; wmw5@cornell.edu

Abstract

Rindermann's analyses reveal substantial correlations for intelligence tests in multiple nations, across content domains and over time. Although impressive and supportive of g-theory, high correlations do not necessarily reflect immutability of g over time or high heritability. Multiple studies demonstrate strong training effects for g-loaded tasks, and tendencies for general intelligence to vary by country may reflect resource and experiential factor differences more than heredity. Copyright © 2007 John Wiley & Sons, Ltd.

The formal theory of general intelligence (*g*; Spearman, 1904) continues to adapt to new methodologies and measurement paradigms, such as rule-space, arpeggio and Bayesian Nets approaches. Indeed, *g*-theory's greatest contributions have been statistical. Rindermann (this issue) advances the measurement and theoretical thrust of *g* with his monumental study; but first, some history is warranted.

Spearman (1904, p. 284) argued that an underlying factor, *g*, permeated all mental activities. He referred to *g* as 'the universal unity of intellective functions . . . that all branches of intellectual activity have in common'. Spearman considered *g* to be mental energy. But others argued that intelligent individuals do better on IQ tests (and other '*g*-loaded' tasks) because their superior central processing mechanism enables them to better detect, store and retrieve important information and relationships from their environments. A core assumption is that this information is available to all individuals except those in the most seriously deprived environments (see Clarizio, 1982; Jensen, 1980).

DOI: 10.1002/per

Psychometric analyses indicate that *g*-loaded test scores (e.g. IQ and major international achievement tests such as PISA, NELS, NAEP) provide the bulk of predictiveness of large aptitude batteries. General intelligence or *g* has been repeatedly shown to provide much better prediction than scores on specific ability measures such as mathematical reasoning, verbal aptitude, perceptual speed or memory (e.g. Humphrey, 1979; Jensen, 1986; Thorndike, 1985).

Critics of *g*-theory have long argued against a single underlying processor that accounts for substantial inter-test variance and is innately determined and immutably fixed (see Ceci, 1996). In his target paper, Rindermann provides numerous convergent analyses that support *g*-theory's predictions that the first unrotated principal component (*g*) is more stable and a better predictor than any specific ability measure (and also, although he does not assess it, more heritable). Rindermann reveals very high correlations between different content measures within the same year of school (mean $r = .96$, corrected: $r = .98$); across years of school, correlations within the same data sets between different scales are extremely high (between $r = .91$ and .97, with correction, see Table 1). This is evidence that achievement in mathematics is highly correlated with achievement in the sciences, verbal abilities and problem-solving skill.

Rindermann's analyses raise interesting questions about the presumed multiplicity of intelligences (Gardner, 2006). The huge magnitude of *g* accounting for over 90% of the variance of the 20 achievement measures and intelligence tests was surprising! As Jensen (1998) and others argue, this implies that the cognitive demands underlying different tests are highly similar if not identical even though the tests appear dissimilar (reading, mathematics, science and writing). Rindermann wisely notes that *g* can be affected by similar developmental challenges across all nations. Thus, *g* need not reflect the extent to which all achievement and aptitude tests are saturated with the same underlying ability or resource pool, such as size of working memory, signal-to-noise ratio in the transmission of information in the nervous system, attention span, speed of nerve conductance or cortical glucose metabolism (e.g. Grabner, Neubauer, & Stern, 2006). As van der Maas, Dolan, Grasman, Wicherts, Huizenga, and Raijmaker (2006) argue, these studies have produced interesting correlations but have not revealed the single underlying cause of the *g*-factor (Ackerman, Beier, & Boyle, 2005; Luciano, Posthuma, Wright, de Geus, Smith, & Geffen, 2005). They employ a non-linear dynamical model to show that psychometric *g* is not incompatible with a large environmental (Flynn) effect.

As Rindermann notes, even different tasks require the same cognitive processes: High correlations between mathematics and reading scores may be due partly to the need to read and comprehend mathematics questions. Specifically, test information has to be encoded, stored and compared with knowledge in memory; this new information must be structured and comprehended, often involving reasoning processes; similarities and differences have to be retrieved. This process often entails abstract thinking, and problem solving on achievement tests often depends on mental speed, concentration, time management, motivation, low test anxiety, metacognitive and test-taking strategies, especially under time pressure, as with the TIMSS.

Do these high correlations signify that *g* must reflect the operation of a singular resource pool in the central nervous system that mediates all cognitive activity? No—high correlations do not require that *g* be heritable or immutable. Recently, Feng, Spence, and Pratt (2007) showed that gender differences in spatial attention, a basic capacity that supports higher level spatial cognition, can be remedied by playing videogames. Remarkably, they show that playing an action videogame for 10 hours virtually eliminated

DOI: 10.1002/per

the gender difference in spatial attention and that this game playing transferred to mental rotation ability, a higher-order cognitive sex difference. Others (e.g. Hansen, Heckman, & Mullen, 2004; Winship & Korenman, 1997) have shown that schooling increases the AFQT score (a form of intelligence) on average between 2 and 4 percentage points (more at the low end), roughly twice as large as the effect claimed previously (Ceci, 1991).

Although these demonstrations of huge environmental effects on intelligence and achievement do not rule out the operation of heritable sex differences, they do illustrate the danger in inferring that *g* reflects immutable differences. The stability Rindermann reports for the TIMSS across different grades is both very high and *g*-loaded (corrected math and science correlations $r = .94$ and .90), but this does not mean that it is ineluctable, as the videogame example illustrates. The high correlations between mathematics and science within a grade ($r = .95$), and the high first-factor loadings of ability test scores across nations may reflect uniformity of experiences and challenges as much as uniformity of ability differences across nations. Different experiences could alter the aptitude and achievement patterns significantly.

In closing, we thank Rindermann for his Herculean effort. His accomplishment required not only a massive amount of work to co-ordinate these data archives, but also something far more important: Rindermann bridges myriad research traditions, specifically psychometrics, education, economics, cognitive science and cultural studies. He has not solved all of the enigmas or completed the synthesis of these different traditions, but he stands out among a handful of others as being bold enough to cast nets this broadly (e.g. Flynn, 2007; Styles, 2006).

A *g*-Factor of International Cognitive-Ability Comparisons: What Else?

FILIP DE FRUYT

Department of Developmental, Personality and Social Psychology, Ghent University, Belgium
Filip.DeFruyt@ugent.be

Abstract

Rindermann showed that student assessment means across countries are strongly correlated with intelligence means. Potential reasons for this strong relationship are discussed, and alternatives for student assessments are considered and evaluated.

Rindermann (this issue) showed that correlations of ability means across countries are very high, within student assessments across ages/grades using the same test, between different student assessments using multiple tests and also between an intelligence measure or estimate and different student assessments. A single factor, entitled a *g*-factor of cognitive ability, explained nearly 95% of all mean cognitive ability variance across countries. Such findings are testimony—at a cross-national level—of what Spearman (1927) called almost a century ago the 'indifference of the indicator principle', referring to *g* as a latent construct considered important and necessary for a wide range of problem-solving tasks and for knowledge acquisition. To the extent that different items enclosed in student

DOI: 10.1002/per

assessments can be solved with the information provided in the tasks themselves, test results will correlate with each other, and also with general mental ability (GMA). Rindermann (this issue) demonstrates that aggregation of individual assessments within different nations ultimately cumulates in meaningful mean-level differences across nations.

At the level of the individual, intelligence or *g* has been shown to be among the most powerful psychological predictors of human performance, including job performance and training proficiency across nations (Salgado, Anderson, Moscoso, Bertua, & De Fruyt, 2003) and different occupations (Salgado, Anderson, Moscoso, Bertua, De Fruyt et al., 2003), but also study performance in primary, secondary and undergraduate and graduate training (Kuncel, Hezlett, & Ones, 2001; Lounsbury, Sundstrom, Loveland, & Gibson, 2003; Spinath, Spinath, Harlaar, & Plomin, 2006). Therefore, it is not surprising to find strong relationships between *g* and student assessment tests, even when measurements are aggregated across individuals within countries, and to observe that differences in intelligence across countries parallel differences in student assessments.

What should attract our attention, however, is the size of this relationship that seems only slightly affected by adopting different corrections to account for sampling inadequacies. Indeed, at the inter-individual level, intelligence measures and school performance criteria are usually correlated in the .20–.60 range, suggesting that these constructs are not substitutes for each other, and that other factors beyond *g*, for example personality traits, contribute to school and academic performance. This discrepancy calls to reflect on the status of such aggregated student performance measures and further requires an examination of the nature of the student assessment tests, their ultimate purpose and use in educational policy making.

Aggregation across different items composing a scale or aggregating scores over many individuals, usually leads to a more reliable assessment of a phenomenon of interest, with superior psychometric properties. The sizeable associations that are observed correlating mean students' assessments across different countries can be probably partly explained as a beneficial side-effect from aggregation, given the substantial mean-level differences for student assessments across countries. However, as Rindermann also argues, the items enclosed in student assessment tests can be almost considered as intelligence items (and vice versa), so the observed strong associations can be hardly surprising, suggesting almost an isomorphism at the level of the operationalisation between intelligence and student assessments, after aggregation. The major question then is whether this a desirable situation?

Student assessment tests are mainly constructed for comparative purposes, i.e. within countries to compare schools and examine school-effectiveness, and across nations to compare educational performance levels and examine return on investment of different educational policies. Considering the findings from the present study, showing that student assessment means across countries are strongly correlated with intelligence means, leads to the question how much these student assessments (and as a consequence also the intelligence means) reflect the impact of factors such as educational policies, or differences in mean personality traits or motivation across countries, beyond genetic factors or the interaction between genetic and environmental influences. Rindermann's findings will probably encourage the call for alternative operationalisations of student assessments.

How is the content of student assessments more recently and alternatively defined and operationalised? Over the past decade, one could notice that the competency movement became very influential in professional fields dealing with individual differences, such as the domains of human resources, but also instruction and education. In instruction and

DOI: 10.1002/per

education, the final objectives or 'end terms' per grade are defined more and more in terms of 'competencies'. Competencies still refer to the acquisition of particular problem-solving strategies, but are also partly knowledge based, or refer to skills, behaviours, attitudes and even enclose values. To the extent that student assessment studies—for comparative or evaluation purposes, within or across nations—have to reflect mastery of competency-oriented 'end terms', their content will have to be considerably amended or changed. Nevertheless, basic individual differences, including intelligence, are underlying building blocks of competencies (De Fruyt, Bockstaele, Taris, & van Hiel, 2006) so there will always be a (sizeable) association with intelligence.

Moreover, the switch to competency-based assessment is a difficult and a highly underestimated endeavour. First, there is almost no consensus on the competency model that should be adopted, although the differences among models are usually not that large (Furnham & Hogan, 2006). Moreover, it is usually suggested that competencies are relatively easy to assess, although there is no evidence at all for such a claim. On the contrary, the multifaceted nature of many competencies makes it very difficult to obtain reliable ratings or assessments, especially when the time for assessment is limited and large groups have to be examined at a time. Without defending student assessments in their present form, those criticising the content, nature and correlates of the student assessments examined by Rindermann, should realise that a switch to competency-based student assessments holds a high risk of finally ending up with student assessments that do not correlate across time and across different tasks, and show no relationship with intelligence, nor at the inter-individual level, nor at the cross-national level of mean comparisons. What else then?

The *g*-Factor Is Alive and Well But What Shall We Do About It?

ANDREAS DEMETRIOU

University of Cyprus and Cyprus University of Technology, Cyprus
ademetriou@ucy.ac.cy

Abstract

This commentary first demonstrates, in agreement with the target paper, that g exists, and it specifies the cognitive dimensions involved in it. It then argues that g is malleable and plastic and specifies how it can be increased. It is also maintained that rank ordering groups along a dimension of g is possible but difficult to achieve and that the present international studies depart from this ideal. Finally, it is argued that we need policies and programmes for the cultivation of g in the best interests of everybody. Copyright © 2007 John Wiley & Sons, Ltd.

This commentary deals with four questions: (1) Does a *g*-factor really exist? (2) What does it involve? (3) Is it malleable? (4) How can *g* be educated? The first two are related to the validity of the positions advanced by Rindermann. The last two are concerned with their implications.

DOI: 10.1002/per

Does a g-factor exist? It certainly does and it came back full force in the recent years thanks to both the integrative work of scholars such as Carroll (1993) and Jensen (1998) and the use of new methods and technology for the study of intelligence, such as structural equation modelling and neuroimaging. Psychometrically speaking, *g* is a higher-order construct reflecting the so-called 'positive manifold'. That is, it reflects the fact that all tests are positively correlated (Carrol, 1993; Jensen, 1998). The more variable the tests, the stronger the *g*-factor is (Humphreys & Stark, 2002). This statistical construct is supposed to reflect the operation of an invisible power, which, like gravity (Detterman, 2002), underlies and constrains performance on all of the tests and is responsible for the positive manifold (Demetriou, Christou, Spanoudis, & Platsidou, 2002).

What does g involve? *g* is specified in reference to five cognitive dimensions: speed of processing, control of processing, representational power, inference and self-awareness self-regulation. Speed of processing reflects the general efficiency for information processing. Control of processing refers to the ability to focus attention on task relevant information, ignore irrelevant information and inhibit irrelevant responses. Representational power refers to the volume and kind of information that can be represented and processed. Inference refers to the processes enabling one to integrate and combine information in order to draw valid and true conclusions not directly obvious in the data. Self-awareness and self-regulation refer to processes underlying self-representation and self-management both on-line and in the long-term in a particular social and cultural context (Demetriou, 2006; Demetriou et al., 2002; Demetriou & Kazi, 2006; Demetriou, Mouyi, & Spanoudis, submitted). Taken together, these processes almost fully exhaust the variance of performance on any cognitive test, such as those used in the target paper.

Recent neuroscience and genetic research provide strong biological support for *g*. Specifically, it is now well established that a number of areas in the frontal and the parietal lobes of the brain are closely associated with general intelligence (Jung and Haier, in press). That is, individual differences in the volume, the organisation and the functioning of these regions are related to individual differences in *g*. In fact, the biological foundations of *g* go deeper than the brain to the genes themselves that control the biological processes underlying the construction and development of the brain. There is now evidence showing that specific genes are responsible for differences in the size and general plasticity of the brain to respond efficiently to information by building the networks required for information processing (Kovas & Plomin, 2006). Also, there is evidence that particular genes are responsible for individual differences in brain networks related to attention and executive control (Posner, Rothbart, & Sheese, 2007).

Is g malleable? It is emphasised that the association of *g* with the brain and the genome by no means implies that *g* is fixed and unchangeable. In fact, genetic research suggests that 50% of variance in intelligence is accounted for by shared genes (Grigorenko, 2002). The rest is left to other forces residing in the environment. Therefore, *g* is definitely plastic and malleable, like the brain itself which carries its operation (Mareschal, Johnson, Sirois, Spratling, Thomas, & Westerman, 2007). The Flynn effect provides strong support to the operation of these forces. According to Flynn (1987), intelligence in the general population rises by about one standard deviation every 30 years. Moreover, cognitive acceleration research suggests that learning focusing on some of the dimensions of *g* above produces stable and transferable increases in *g* (Klauer, 1997; Shayer & Adey, 2002).

How can g be educated? In a recent paper (Adey, Csapo, Demetriou, Hautamaki, & Shayer, in press), we suggested that a programme for educating *g* must satisfy the following requirements:

DOI: 10.1002/per

1. Learning activities must have the potential to create challenge and enable one to deal with novelty.
2. Learning should be collaborative in the sense that learners learn to listen, argue, justify and become accustomed to changing positions and knowledge.
3. Learning must raise awareness of what may be abstracted from any particular domain-specific learning, connecting the present concept to others already in possession, and of the thinking and learning processes as such.
4. Learning must build strategies for handling weaknesses and limitations of one's own processing, representational and computational capacities and capitalising on one's own strengths.

Unfortunately, education nowadays only indirectly and unsystematically addresses these skills and strategies.

Conclusions. The evidence reviewed here suggests that there should be no surprise for Rindermann's findings. Being like gravity, *g* is everywhere and, therefore, its effects can be noticed if systematically looked for. The international cognitive-ability comparisons, despite their possible shortcomings and weakness, are systematic enough to be able to uncover it. In fact, Rindermann is to be applauded for demonstrating the presence of *g* at the national and the international level.

One might then ask: Can social groups and nations be ranked along a dimension of *g*? Ideally, this is possible. Once we have ensured full equivalence of tests, testing procedures and conditions and test-taking skills and experience, the rank ordering of groups would say something about their cognitive potential. However, these ideal conditions are in practice difficult to achieve, if possible at all. In fact, the studies used in the target paper do not seem to fully satisfy all of these requirements. Therefore, any rank ordering that comes out of them must be taken with caution. However, these studies, together with Rindermann's paper, do send a strong message to educators and policy makers influencing education at the national and the international level: We must develop programmes and methods that would raise *g* in both the individual and the group at all levels, from local to global, mobilise the necessary resources, and of course refine and sharpen our diagnostic tools. I believe that going in this direction is a move away from the reasons that cause poverty, misery and war within and between nations.

What Lies Behind *g*(I) and *g*(ID)

JAMES R. FLYNN

Emeritus Professor, Political Studies, University of Otago, New Zealand
jim.flynn@stonebow.otago.ac.nz

Abstract

Rindermann's results suggest that different factors lie behind the emergence of g in international comparisons and the emergence of g when we compare the differential performance of individuals. This renders g(I) and g(ID) so unlike that they have little significance in common. Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/per

Rindermann (this issue) uses Piaget to analyse data derived from conventional mental tests. Although a new convert to this approach, I am convinced of its merits. In my recent book *What is Intelligence*? I found it indispensable in accounting for the enormous size of IQ gains since 1900 in America (Flynn, 2007).

Rindermann analyses international comparisons on a variety of cognitive measures, ranging from IQ, through competence in mathematics, through competence in science, to reading literacy, and finds remarkable uniformity. There is little variation in a nation's ranking in terms of what cognitive measure is used. He therefore derives a *g* close to the limit of 1.00 and *g* loadings for the various measures of the same magnitude. In accord with the developmental tradition, he gives a functional analysis of what lies behind this *g*, a practice often neglected by those who tend to take the latent traits produced by factor analysis as the keys to explanation. He notes that various nations are at different points on the road to modernity and suggests that this is an important factor that lies behind the pattern of the data.

I concur but want to perform two tasks: (1) Compare Rindermann's 'international *g*' with the 'individual differences *g*' factor analysis yields, when it is applied to data comparing subjects all of whom belong to the same time and place; (2) Illuminate the significance of the gap between these two kinds of *g* by functional analysis. By functional analysis, I mean explaining what lies behind factor analysis in terms of physiology, social conditions and the cognitive behaviour rooted in those social conditions. I will use *g*(I) and *g*(ID) to label the two kinds of *g*.

Let us remind ourselves why *g* exits at all. An entity whose score exceeds the average performance on one cognitive measure tends to exceed the average performance on a range of other cognitive measures. This creates a positive manifold with a large first major factor. For example, on the 10 subtests of the Wechsler Intelligence Scale for Children (WISC), individuals who do above average on the vocabulary subtest tend to do better on the other nine. We can render *g*(ID) more sophisticated by distinguishing tests where an acute mind is actively at work solving problems, fluid *g* of the sort attached to Raven's Progressive Matrices, from those tests that measure the kind of accomplishment that an acute mind would be likely to attain throughout life, the crystallised *g* of vocabulary. And we find secondary patterns of superior performance: cases in which superior individuals tend to do better on visual tasks than verbal ones, giving rise to visual, verbal, quantitative factors and so forth.

The most impressive thing about *g*(ID) is that it tends to rise with the cognitive complexity of the task (Raven's) or the cognitive complexity of the mental processes by which cognitive accomplishments are acquired (developing a large vocabulary). Digit span backward sets a more complex task than digit span forward and its *g* loading is higher, which is to say that individuals who are above average tend to better the average person by a greater gap on the former than the latter. This consistent superiority across complex tasks has engendered the hypothesis that high-IQ subjects have a better brain physiology than low-IQ subjects. This may be true to some degree, for example, they could be closer to the optimum blood supply to the prefrontal lobes or their dopamine may operate better, that is, strengthen synapses more efficiently when they are activated by cognitive exercise (Flynn, 2007, Chapter 3).

So what lies behind *g*(ID) is a correlation between *g* loadings and cognitive complexity, perhaps a physiological superiority of the some people's brains over others, and a range of subordinate factors. What lies behind *g*(I) is so different that none of these need apply. When the entities being compared are nations separated by degrees of modernity, rather than individuals who profit from whatever degree of modernity prevails at their time and

DOI: 10.1002/per

place, there is no reason to assume an average physiological superiority of one nation over another. Or better, if that exists, it is a by-product of modernity in that modernity is productive of better food supply. And there is no reason to expect *g*-loadings to attend degree of cognitive complexity.

I suspect that there are differences of cognitive complexity in the measures Rindermann uses. But if modernity is equally potent in boosting IQ and school achievement, then that would act as a leveller that conceals such differences behind uniform and virtually perfect *g* loadings. Since modernity sets in action a process by which rising formal education and rising IQ reinforce one another by reciprocal causality, between-nations comparison would blur distinctions between their measures. It would equally inhibit the appearance of subordinate factors. If all cognitive skills ranging from Raven's to quantitative to verbal to visual are at the mercy of modernity, they will rise in tandem.

My comments are a caution, which Rindermann does not need, against thinking that all *g*s have the same significance. Lynn and Vanhanen (2002) interpret the varying mean IQs of nations as evidence of differences in physiological quality with a genetic component. Whether this is true or not, the presence of an almost perfect *g*(I) adds no support. Its significance may be totally different to the presence of *g*(ID). I suspect that the 21st century will see a levelling of between-nation IQ differences as developing nations travel the road to modernity that developed nations travelled in the 20th. There are signs that the day of massive IQs rises may be ending in the developed world (Schneider, 2006) and just taking off in nations like Kenya (Daley, Whaley, Sigman, Espinosa, & Neumann, 2003) and Dominica (Meisenberg, Lawless, Lambert, & Newton, 2005).

If that trend continues, the international gaps will tend to disappear and we will know, rather than merely conjecture, than modernity is the glue of the present *g*(I) and not varying intelligence or brain physiology.

Parsimony or Reductionism?—Against the *g*-Factor of Nations

HEDE HELFRICH

Department of Psychology, University of Hildesheim, Germany
helfrich@uni-hildesheim.de

Abstract

Rindermann presents the thesis that international student assessment studies primarily measure the same cognitive ability as intelligence tests, namely Spearman's g-factor. My comment focuses on two objections to his analysis. First, the uniform correlations taking countries as units of analysis may mask different shapes of correlations within each country. Second, the psychometric approach to assess national differences in general intelligence cannot claim universal validity. Copyright © 2007 John Wiley & Sons, Ltd.

The central thesis of Rindermann's provocative paper (this issue) is that international student assessment studies primarily measure—at the macro-social level—the same

DOI: 10.1002/per

cognitive ability as intelligence tests. This ability can best be described as Spearman's *g*-factor (Jensen, 1998; Spearman, 1904). According to the author, the causes of intelligence, knowledge and performance are similar and nested in reciprocal causation. Genetic factors, family background and school education simultaneously affect this common cognitive ability in different achievement and intelligence tasks.

As an implicit consequence, we may conclude that student assessment studies such as PISA or TIMSS can indicate the cognitive level of a country but are not at all suitable to indicate the quality of schooling in a country.

In supporting his thesis, Rindermann primarily relies on correlations between different student assessment tests as well as on correlations between the combined assessment measures and intelligence tests at the level of national data. Evidence of the unidimensionality of cognitive abilities is provided by a factor analysis of 20 student assessment scales and the intelligence test collection of Lynn and Vanhanen (2002, 2006), yielding a strong *g*-factor of differences between nations.

The issue Rindermann's paper addresses is a very fundamental one, both in the study of cognitive abilities and in cross-national comparisons. My comment will focus on the following question: Do the cross-national correlations between different scales justify the assumption of a universal *g*-factor? In my view, Rindermann's analysis poses two serious problems that may confound the results.

1. The uniform correlations taking countries as units of analysis may mask different shapes of correlations within each country. There is no logical basis for inferences about intra-group variation on the basis of inter-group variation. This especially holds for the correlations of the student assessment sum with intelligence tests ($r = .85$ and $r = .86$). Imagine the fictive example of three different countries *A*, *B* and *C* given in Figure 2. Let us assume that all three countries have equal means and equal standard deviations in the intelligence scale ($M = 100$, $SD = 15$) as well as in the student assessment scales ($M = 500$, $SD = 82$). However, as Figure 2 illustrates, different shapes of correlations may be found within each country: Country *A* may show a strong positive correlation (consistent with the correlation at the macro-level), country *B* may show a strong negative correlation (inconsistent with the correlation at the macro-level) and country *C* may show a curvilinear relationship ($r_{\text{linear}} = .58$; also inconsistent with the correlation at the macro-level). Case B may be considered implausible, but case C is not unrealistic (meaning that performance is best given a medium level of intelligence). There is some empirical evidence of differences between IQ and achievement within different cultural groups. For example, Flynn (1991) proposed that students from Asian cultural backgrounds typically achieve at higher levels than non-Asian students with the same IQs. Consistent with Flynn's hypothesis, in a study on relationships between IQ (measured with Raven Progressive Matrices) and academic achievement among Australian school children, it was found that students from Asian backgrounds obtained higher mathematics grades than their Anglo-Celtic Australian peers with the same IQ (Dandy & Nettelbeck, 2002).

To a somewhat lesser degree, the argument that correlations at the macro-level may mask correlations at the micro-level also holds for the correlations within the student assessment studies. Although Rindermann refers to substantial correlations at the individual level within the different tasks of PISA ($r = .65$ and $r = .71$) and within the different tasks of TIMSS ($r = .61$) in the German sample, no correlations at the individual level are presented between PISA and TIMSS scales (see also Rindermann, 2006). Against the case of homogeneity, the author himself concedes 'some statistical support for *content-specific differences* . . . given by TIMSS and PISA'. Moreover, he presents evidence for

DOI: 10.1002/per

culture-specific intra-group correlations when he states that, in Israel, there are higher values in reading literacy than in mathematics and science within PISA.

The intelligence test data collection of Lynn and Vanhanen (2002, 2006), which was intended to measure general intelligence ('*g*') and to represent it on one common standard scale (with mean = 100 and $s = 15$), cannot claim psychometric cross-cultural equivalence. Apart from the fact that, in comparison to the student assessment studies, the representativeness and comparability of the samples is low, there are fundamental objections to a psychometric approach whose goal is the assessment of cross-cultural differences in general intelligence (see Helfrich, 1999, 2006). If 'general intelligence' sensu Spearman's *g* is to be measured, it is (unlike PISA and TIMSS tasks) not the content of individual test tasks which operationalises the concept but a representative sample of different tasks taken from a hypothetical multitude of possible tasks (see Cronbach, 1984). That means that the various types of tasks are so conceived that, taken as a whole, they operationalise the construct 'intelligence'. If the combined scores of this ensemble were to be compared on a common metric scale valid for all cultures, one must, in addition,

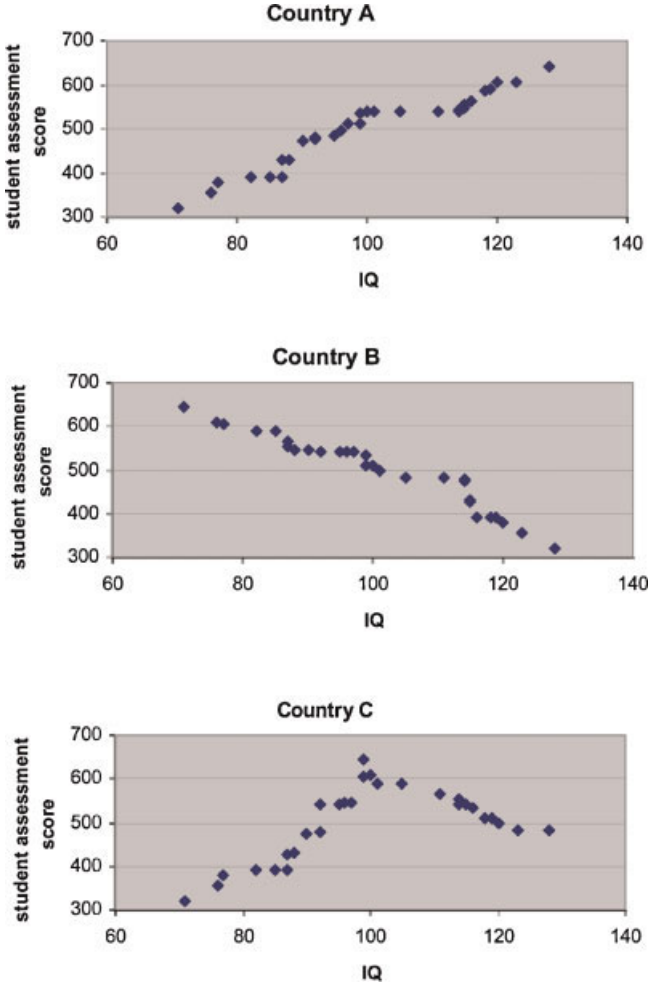

Figure 2. Correlations between IQ and student assessment score in three fictitious countries.

DOI: 10.1002/per

guarantee that the structure of the correlations between individual tasks is similar in all cultural groups investigated. There is empirical evidence that representativeness and structural equivalence cannot be simultaneously attained in cross-cultural research (see Poortinga & van de Flier, 1988; van de Vijver & Leung, 1997). Thus, one cannot conclude from the observed performance that the competence structure is comparable. Theoretically, the lack of generalisability and the lack of structural equivalence must be attributed to the culture-specific transformation of situational demands into cognitive tasks. What is considered as 'intelligent' refers to the successful adaptation to those cognitive tasks which are significant within a specific culture. As Sternberg (1988) in his 'triarchic view of intelligence' suggests, there may exist universal *components* of intelligence. However, their weights in contributing to general intelligence may vary due to different environmental *contexts* and due to different *experiences*.

Despite my criticisms, I very much appreciate Rindermann's fruitful endeavour to bridge a gap between different research traditions which are not only isolated from another, but sometimes completely ignore each other. Future research will profit from his analysis.

Contributions Made and Contributions Yet to be Made by Rindermann's Analysis of National IQs

EARL HUNT

Department of Psychology, The University of Washington, USA
ehunt@u.washington.edu

Abstract

Rindermann presented criteria for judging international assessment studies of cognitive competence. He showed that broad trends in the present data are sufficiently strong so that the weaknesses of some data sets can be disregarded. We still need to understand how the nature of the testing situation itself restricts the measurement of important aspects of cognition, at the national level. Copyright © 2007 John Wiley & Sons, Ltd.

In the last few years several studies have reported measures of 'national IQ', scholastic ability, reading, mathematics accomplishment, or some other such term. It would probably be better to say 'National Cognitive Competence'. Differences have been shown between nations, and speculations about causes and implications have been made.

This is a serious issue. Human capital, the level of cognitive competence in the population, is an important resource. Consider Japan and South Korea. In less than half a century both countries rose from the devastation of war to become two of the most prosperous countries on the globe, virtually entirely on the ingenuity of their people, for neither country is rich in natural resources. Other countries seem stuck, in spite of inputs of wealth and, in some cases, abundant natural resources. Economists and politicians point to a variety of relevant factors: free market economy, political and social stability, the rule of law and democratic practices. Economists and politicians have, however, generally not

DOI: 10.1002/per

looked at human cognition; intelligence for short. There seems to be an implicit belief that there are no national-level differences in intelligence, but this may very well not be the case. In fact, the studies of national intelligence challenge this belief.

There are two reasons that national-level differences in intelligence have been disregarded. One is that it can be argued that intelligence, as evaluated by these tests, is a Western concept, and that the abilities evaluated by the tests may not be the ones valued by non-western societies. This is a spurious argument for two reasons. First, the economic indicators we are trying to relate to intelligence are also Western concepts. As the commentator Thomas Friedman has said, the world is flat. We are not asking whether or not various national populations have the ability to compete in their own societies, we are asking about their ability to compete in the Western-defined international marketplace. The tests are appropriately designed to address this question.

The second reason that some people disregard the tests is that they can be attacked as invalid for a variety of weaknesses, ranging from downright errors in reporting (a major problem in the IQ data) to non-representativeness of examinees in the scholastic assessment data. Rindermann has made a major contribution by showing that these weaknesses, although they do exist, are probably not enough to affect the outcome. He showed that even if you make substantial corrections for various artefacts, the correlations between scores are very little affected.

Rindermann's next contribution was to show that the tests are highly correlated, so that one can think of a general level of cognitive competence that varies across nations. This is in agreement with earlier work by several authors; Rindermann's analyses is a welcome support for the views of people like Wittmann and Lynn and Vanhanen in this regard. Rindermann then raises the question of why this should be so. He points to common test material as an obvious reason. Although the tests differ in their names his informal cognitive task analysis of the tests shows that they all draw on some common cognitive functions. This alone would be sufficient to produce a single factor in the international data, regardless of whether or not one believes in general intelligence at the individual level.

Let me suggest something that Rindermann does not consider. All the tests are given in 'drop in from the sky' format; examinees are given a series of problems to be completed in a relatively brief period of time, quite removed from the context of everyday living. Tests of this sort can and do tap some important aspects of cognition, otherwise they would not work at either the individual or national level. What they do not tap are traits of reflectivity, persistence, personal discipline and characteristic reactions to frustration. Recent work in the educational field suggests that these are important moderators, not of test scores, but of accomplishment over a substantial period of time. That sort of accomplishment, not the ability to score well on a timed test, is what drives economic and social development. I am not saying that the abilities tapped by the tests are unimportant, they are important. But they are not the whole story. A full understanding of the importance of human capital will require understanding reflectivity, persistence and personal discipline as well as understanding the narrower sorts of intelligence evaluated by a 1–3 hour test.

Finally, we have the chicken and egg problem; do national levels of intelligence, broadly defined, drive economic and social development, or is it the other way around? Virtually everyone believes the relations are reciprocal. Understanding the reciprocal pathways is not impossible, providing that we take 20 or 30 years to gather the necessary longitudinal data. Do psychologists have the patience that workers in fields like climatology and

DOI: 10.1002/per

geology do? For that matter, is it possible to study social systems for long periods of time without having the system under study perturbed by exogenous events, such as wars, global recessions or outbreaks of pandemics? I hope that the answer to both of these questions is 'yes', for the issues raised by observations of differences in national intelligence can only be answered by such research.

IQ and Inequality in Human Conditions: Are Correlates Dependent on the Level of Analysis?

PAUL IRWING

Manchester Business School, University of Manchester, UK
paul.irwing@mbs.ac.uk

Abstract

Rindermann greatly increases the credibility of Lynn and Vanhanen's (2002) findings, by reproducing those using international studies of student attainment. Additional candidates for the prediction of economic outcomes include: personality, knowledge, motivation, psychopathology, inter-group processes and efficient employment of human resources, the importance of each being dependent on whether analysis is at the individual or cross-national level. Copyright © 2007 John Wiley & Sons, Ltd.

In 2002, Lynn and Vanhanen published a ground breaking book, which demonstrated that economic variables measured at national-level correlate substantially with the psychological variable of general cognitive ability (IQ). Given the reasonable inference that productivity is related to income generation, this finding reproduced, at the national-level, the well-established individual-level correlation between IQ and job performance estimated at .51 in American (Schmidt & Hunter, 1998) and .62 in European studies (Salgado et al., 2003). Lynn and Vanhanen (2002) observed correlations, which varied in magnitude from .72 to .78, for IQ with Gross National Product measured at purchasing power parity, probably their best measure of national economic performance. Subsequently, it was shown that this finding generalises to a range of indices of the human condition (Lynn & Vanhanen, 2006). Criticism of these studies highlighted that sampling adequacy and comparability of IQ measures was low (Hunt & Wittmann, in press). The major contribution of Rindermann's paper is to show that international studies of student attainment, in which sampling adequacy and comparability of measures is generally good (1) measure essentially the same construct as Lynn and Vanhanen's measures of IQ, and (2) evidence similar relationships to economic indicators both separately, and when an aggregate index of *g* is derived from studies of educational attainment combined with the IQ measures of Lynn and Vanhanen.

However, IQ does not perfectly predict economic outcomes, so the question arises as to what variables are related to the unexplained variance. This question has been addressed by Lynn and Vanhanen (2002, 2006) in terms of a largely economic analysis.

DOI: 10.1002/per

Given that the individual-level relation between IQ and job performance is reproduced at the cross-national level, it seems that other psychological variables related to economic performance should also be prime candidates as explanatory variables. The meta-analysis of Barrick and Mount (1991) has established that the Big Five factors of personality (Extraversion, Emotional Stability, Conscientiousness, Agreeableness and Openness to Experience) all show generalisable, if small relationships to job performance, and that the magnitude of the average correlation is increased to about .40 when prediction is based on a job analysis (Tett, Jackson, & Rothstein, 1991). Further, the mean correlation of IQ with job performance is increased from .51 to .65, when an integrity (personality) test is added (Schmidt & Hunter, 1998). Knowledge has also been found to relate to superior job performance (Ericsson & Kintsch, 1995), and in a laboratory task to predict performance more strongly than working memory (Hambrick & Engle, 2002). The latter suggests that knowledge may be a more important predictor of job performance than is IQ, given the admittedly contested hypothesis (Ackerman et al., 2005), that working memory and IQ are synonymous (Kyllonen, 2002). In discussions of the factors which predict job performance, motivational factors have also featured consistently (McCloy, Campbell, & Cudeck, 1994; Smith, 1994). An individual variable not normally considered in this context is psychopathology, yet this must be a serious candidate given that the related variable of job stress has been shown to cost the UK economy billions of pounds.

A complete explanation should probably also consider group and societal level variables such as inter-group processes (Haslam, 2001) and fair employment. The work of both Rindermann (this issue) and Lynn and Vanhanen (2006) highlights the incidence of violent conflict as an explanation of the underperformance of some economies, however, the failure rate of mergers and acquisitions is extremely high, and a plausible explanation of these also lies in inter-group dynamics. With regard to fair employment, utility theory has shown that optimal methods of personnel selection could substantially improve the performance of most economies (Boudreau, 1992; Schmidt & Hunter, 1998).

Although individual and cross-national level data clearly exhibit similar effects, there is an issue of the relationship between analyses at these different levels. For example, inefficient allocation of talent to appropriate positions probably grossly reduces within-nation relationships between IQ and economic outcomes. Indirect evidence of this is provided by the typical estimate of the correlation between IQ and income at .30, much lower than would be expected on the basis of the correlation between IQ and job proficiency. However, if all societies are similarly inefficient in their allocation of human resources to jobs, at the level of national economic performance, this source of variability would be randomised out, producing the observed stronger cross-national relationship between IQ and economic performance. Conversely, assuming that within-nation variability in personality is greater than that between nations (Irwing & Bedwell, 2006), then correlations of personality with economic indicators at the cross-national level may be smaller than the already tiny correlations found at the individual level.

From the perspective of Rindermann's paper, a particularly important manifestation of the variations attributable to different levels of analysis may be exemplified by the relationship between educational attainment and IQ. As we have already observed, Rindermann's confirmatory factor analysis shows that educational attainment and IQ are virtually synonymous at the cross-national level, as evidenced by the factor loading of IQ on the general factor at .96 being virtually identical to the average loading of educational

DOI: 10.1002/per

attainment tests at .99. This compares with arguably the best estimate of the individual-level correlation between educational attainment and IQ at .80 (Deary, Strand, Smith, & Fernandes, 2007). What accounts for this discrepancy? At the individual level, the two principal distinguishing features between educational attainment and IQ are probably that the former tests mastery of a curriculum, as opposed to setting novel problems, and secondly, that the forms of assessment differ with short multiple choice items typical of IQ tests, and extended essays or reports more commonly found in educational settings. However, these differences do not necessarily obtain, and the cross-national assessments of educational attainment usually remove both these features (e.g. PISA). This is because, for these comparisons to be valid, it is necessary to control for differences in curriculum across countries, while multiple choice items are required to ensure measurement invariance (Meredith, 1993). Hence, at the international level, educational attainment and IQ are effectively isomorphic.

It Does not Help to Ignore It

WENDY JOHNSON[1,2]

[1]*Department of Psychology, University of Edinburgh, UK*
[2]*Department of Psychology, University of Minnesota, Twin Cities, USA*
wendy.johnson@ed.ac.uk

Abstract

Rindermann's (this issue) analysis highlights the large factor of cross-national differences that closely links student achievement and intelligence test results. For some, existence of this factor justifies their dominance by proving their innate superiority. For others, it is an awkward observation to be buried amid hope that it will disappear by itself. Neither is constructive; neither is likely to be accurate. Copyright © 2007 John Wiley & Sons, Ltd.

It is easy to summarise Rindermann's analysis by claiming that, when aggregated at the national level, student achievement tests assess national general intelligence, with all the implications of immutable innate capacity that brings. There is certainly evidence that, within nations, student achievement and intelligence test scores are strongly correlated. There is also evidence that there are substantive genetic influences common to both forms of test scores within nations. Moreover, there is nothing inevitable about Rindermann's result across nations in principle. That is, it is possible in principle that Rindermann's analysis could have revealed independent or even negatively correlated factors of student achievement and general intelligence. Thus Rindermann's result is important because it tells us something about how our world is working, something that is better acknowledged and discussed than buried amid hopes that, like a misbehaving child, it will stop if we just ignore it.

For Rindermann's analysis to have come out differently, achievement and intelligence would have had to have been impacted independently or in opposite ways and more strongly by nationhood than by the forces that link the two within nations. That is,

DOI: 10.1002/per

something about nationhood would have had, for example, to act strongly to increase intelligence test scores but decrease achievement test scores, or at least to separate the two. The fact that such situations can occur, in which the sources of individual differences within groups are different than the sources of differences between groups, has long been recognised by behaviour geneticists and by some personality and cross-cultural psychologists. It is difficult to conceive of what the forces of nationhood might be in the achievement and intelligence situation that could act in this way. This difficulty contributes to the temptation to conclude that Rindermann's result says something about how the world *must* work always beyond how it *does* work currently. We have evidence that strong environmental forces that could cause such effects do exist, however, in the form of the well-documented and steady increase in intelligence scores over the past 100 years or so known as the Flynn effect.

Among other things, the way the world *does* work currently is an outgrowth of historical and current displays of colonial and military power and their effects on both past and present distributions of wealth. At the same time, despite Rindermann's correlations, the differences in intelligence and student achievement between nations are small in relation to the differences within nations. Moreover, there is substantial evidence for direct effects of environment. Studies of children adopted from deprived circumstances into more favourable ones show that, though their intelligence scores may remain more highly correlated with those of their biological than with those of their adoptive parents, their mean scores are higher than those of their siblings who remain in the deprived circumstances. Immigrants to the United States over the past 200 years also provide substantial evidence that change in circumstances brings change in manifested intelligence: groups have arrived in waves repeatedly from many parts of the world. The first generation arrives 'dumb', displaying great ignorance and poor test scores, but the second generation blends in and acts and tests indistinguishably from the rest of the population. We have choices about educational, political and economic intervention, and assuming that the way the world *does* work currently is the way it *must* work always could be a huge mistake in human terms.

On the other hand, the great social intervention programmes such as Head Start in the United States that were launched with the purpose of increasing the IQ's of disadvantaged youth are generally accepted to have failed in this regard (though they may have succeeded in some others, including improving educational attainment). At present we really do not know to what degree the situation that Rindermann's result documents can be changed. But we do know that high tested intelligence is not necessary to carry out violence and terrorism of a kind that could destroy the wealth and dominance enjoyed by the nations that currently show the highest test scores. And we might suspect that, if they continue to be shut out of the wealth, people in the countries that currently show the lowest test scores will see such violence and terrorism as justifiable options. Whether we do so because we believe it documents our innate superiority or because we hope it will go away by itself, we fail to act upon Rindermann's misbehaving result at our peril.

ACKNOWLEDGEMENT

I thank Bob Krueger for reading over an earlier draft of this commentary.

DOI: 10.1002/per

The Evolutionary Biology of National Differences in Intelligence

RICHARD LYNN
University of Ulster, Northern Ireland
Lynnr540@aol.com

Abstract

Rindermann's work raises the question of the causes of national differences in intelligence. It is proposed that these are likely adaptations that evolved in the European and East Asian peoples to the cognitive demands of survival during the winter and spring in the temperate and cold climates. Copyright © 2007 John Wiley & Sons, Ltd.

Rindermann has done a great job in synthesising data on national differences in cognitive ability from studies of intelligence and attainment in math, science and reading. The correlations across nations between these different measures (at around .80 and .90) are remarkably high, and his finding that a strong general factor is present is important. He is undoubtedly correct in interpreting this general factor as intelligence.

The discovery that there are large differences in intelligence between nations has great potential for explaining differences in a range of economic and sociological phenomena. Developing the exploratory work that I did with Tatu Vanhanen (Lynn and Vanhanen, 2002, 2006), Rindermann calculates that national differences in intelligence are correlated at .61 with *per capita* income, and hence that national differences in intelligence can explain a substantial portion of the problem of why some countries are so rich while others are so poor. Hitherto, nearly all economists have assumed that the peoples of all nations have the same intelligence. Thus, only 7 years ago, Eric Hanushek of the Hoover Institution and Dennis Kimbo of the American National Bureau of Economic Research wrote in the *American Economic Review* that 'we assume that the international level of average ability of students does not vary across nations' (Hanushek & Kimbo, 2000, p. 1191). Rindermann's paper establishes definitively that this is wrong. It should no longer be possible to write this in one of the world's leading economics journals.

Rindermann points out that for far too long the social sciences have been compartmentalised. Since the 1960s' educationists have been reporting that there are large national differences in educational attainment without understanding (or daring to mention) that these are likely to be largely determined by differences in intelligence. Economists use the concept of 'human capital' to explain national differences in economic growth and national wealth without understanding (or daring to mention) that 'human capital' is largely intelligence. Sociologists use the concept of social class to explain numerous phenomena without understanding (or daring to mention) that social class is largely determined by intelligence (Nettle, 2003).

We have now reached a paradigm shift in the social sciences in which intelligence has the promise to become a crucial explanatory and unifying concept, analogous to gravity and energy in the physical sciences. Some progress has already been made in this direction. Sociologists (Weede & Kämpf, 2002) and economists (Jones & Schneider,

DOI: 10.1002/per

2006; Ram, 2007) have confirmed that national differences in intelligence can explain differences in economic growth and *per capita* income. Templer and Arikawa (2006) have shown that national differences in intelligence can explain about two thirds of the variance across nations in homicide rates ($r = -.82$). Kanazawa (2006) has shown that national differences in intelligence can explain about half of the variance across nations in life expectancy (approximately $r = .70$). Meisenberg (2004) has shown that national differences in intelligence are negatively related to corruption ($r = -.68$). Voracek (2005a) has shown that national differences in intelligence are positively related to suicide rates.

Rindermann does not explore the question of what factors are responsible for national differences in intelligence. It is obvious from the maps he provides that the world distribution of intelligence is strongly related to race. It is the European and East Asian peoples who have the high IQs and the African peoples who have the low IQs, while the South Asian, North African and Native American Indian peoples of most of Latin America fall somewhere in the middle. Are these race differences genetic or environmental or some mix of the two?

In attempting to solve this problem I have proposed a theory drawn from evolutionary biology. This theory, originally set out in Lynn (1991) and more recently elaborated in Lynn (2006), is that the high intelligence of the European and East Asian peoples evolved as adaptations to the cognitive demands of survival during the winter and spring in the temperate and cold climates of Europe and North East Asia. Humans evolved from apes and intermediate 'missing link' species in equatorial sub-Saharan East Africa during the last 4 million years or so. These first humans subsisted largely on plant foods, eggs and insects that are available throughout the year. Around 100 000 years ago some groups began to migrate from equatorial Africa into North Africa and the Middle East and then into Europe and the rest of Asia. Between approximately 28 000 and 10 000 years ago these peoples lived through the last Ice Age in which Europe and North East Asia experienced the cold environment of present day Siberia and Northern Canada. Plant foods were not available for a number of months in the winter and spring and so these peoples had to hunt large animals such as deer and mammoth for food. This was more cognitively demanding than gathering plant foods, and the European and North East Asian peoples evolved larger brain size and greater intelligence to solve these problems. In addition, they had to deal with the problems of storing food for future consumption and keeping keep warm, which required the building of shelters and making fires and clothing.

The peoples of North Africa and the Middle East also experienced these problems but to a lesser extent, and so they evolved a level of intelligence intermediate between the European and North East Asian peoples on one hand, and the sub-Saharan Africans on the other. If this theory is correct, we should expect that the coldness of winter temperatures would be correlated with national IQs. Templer and Arikawa (2006) have shown that this is correct ($r = .63$).

DOI: 10.1002/per

If All Tests Measure the Same Thing, How Can We Evaluate the Quality of Schooling?

GERHARD MEISENBERG

Department of Biochemistry, Ross University, School of Medicine, Dominica
GMeisenberg@rossmed.edu.dm

Abstract

The near-equivalence of school achievement tests and intelligence tests raises questions about the causal relationships between intelligence and school achievement, and about the evaluation of educational systems on the background of country differences in general intelligence. Causal relationships remain unresolved, but we can conclude that tests of school achievement can only be evaluated conjointly with tests of fluid intelligence.

How close is the relationship between IQ and school achievement? High correlations between the results of different school assessment programmes (e.g. TIMSS and PISA) and between different school subjects (e.g. TIMSS mathematics and TIMSS science) are not surprising. We have to expect these results if the assessments were properly conducted and if the performance in different academic subjects depends on the effectiveness of the national school systems. More interesting is the observation of high correlations between school assessments and tests of general intelligence. Even at the individual level, correlations of .50–.70 between IQ and school achievement are typical in the absence of range restriction (Jensen, 1998; Mackintosh, 1998), and correlations of .80 are not unheard-of (Deary et al., 2007). The results reported in Rindermann's target paper show that even higher correlations are seen in comparisons between countries.

These results are confirmed by work done elsewhere. In one recent study, the correlation of a summary score of the TIMSS and PISA assessments in mathematics and science with the national IQs reported by Lynn and Vanhanen (2006) was as high as .92 for those 57 countries for which both measured IQ and school achievement are available (Lynn, Meisenberg, Mikk, & Williams, 2007). This is remarkable because few countries at low levels of economic and cognitive development participated in the school assessment programmes. With a larger number of less developed countries, the correlations would almost certainly be even higher. Therefore intelligence tests and tests of school achievement appear to measure the same construct.

What are the causal paths? This raises the question about the relationship between intelligence and school achievement. One theoretical position holds that intelligence is by and large a consequence of schooling. The cumulative evidence of a century of educational research does indeed show that formal schooling raises the performance on intelligence tests (Ceci, 1991). According to one estimate, in the United States each year of additional schooling during the teenage years raises young adult IQ by 2–4 points (Winship & Korenman, 1997). If IQ differences between countries are caused by differences in the quality or quantity of schooling, then we can only expect high correlations between IQ tests and tests of school achievement.

DOI: 10.1002/per

However, there are large IQ differences between children in the same schools, and IQ and school achievement are influenced by genetic factors to about the same extent (Alarcón, Knopik, & DeFries, 2000; Kovas, Harlaar, Petrill, & Plomin, 2005; Wainwright, Wright, Geffen, Luciano, & Martin, 2005). Lynn (2006) proposes that IQ differences between countries are mainly genetic in origin, and that school achievement correlates highly with IQ because both are determined by genetically based differences in intelligence.

There are three possible causes for the covariance of IQ and school achievement across countries: differences in national school systems, differences in the non-school environment and genetic differences between human populations. All three possibilities are theoretically plausible and are at least somewhat supported by empiric evidence. For the set of 123 countries for which all data are available, the correlation of the average IQ in the country with the logarithm of GDP is .81, with the average length of formal education .79, and with skin colour .89. These correlations suggest that all three are important: schooling, non-school environmental factors associated with GDP and the genetic background of the population. Of course, the direction of causality is not always clear.

How can we evaluate school systems? There is, however, one difference between school achievement and IQ: Relative to within-country standard deviations, the between-country standard deviation is higher for school assessments than for IQ. Students in countries with high IQ and high school achievement systematically overperform in the school assessments *relative to* their performance on IQ tests (Lynn et al., 2007). Intelligence itself, rather than GDP, is the important factor for this overperformance. Most likely student's performance depends not only on the student's intelligence but also on the teacher's, and this leads to an incremental effect of intelligence on school performance.

The practical question is: If school achievement is so closely related to IQ, and IQ is determined both by schooling and by other factors, how can we use international school assessments to measure the effectiveness of national school systems? The answer is, of course, that this is possible only by comparing students' performance on the school assessments with their performance on intelligence tests with school-distant materials. The IQs collected by Lynn and Vanhanen (2006) are sub-optimal for this purpose because they are of uneven quality, and they were obtained at widely differing points in time, with different tests and with subjects of different ages in different countries.

If an evaluation of the effectiveness of national school systems is desired, it is therefore mandatory to include a test of fluid intelligence in the international school assessment programmes, with tasks that are unrelated to the specific knowledge and cognitive skills that are taught in school. In this case the students' performance in the school-related subjects can be compared with their general fluid intelligence. The effectiveness of the school system would be indexed not by the students' absolute level of performance in the school assessments, but by their over- or underperformance in the school-related subjects relative to their fluid intelligence.

DOI: 10.1002/per

Does the Globalisation of Assessment Lead to the Globalisation of Education?

FROSSO MOTTI-STEFANIDI

Department of Psychology, University of Athens, Greece
frmotti@psych.uoa.gr

Abstract

International educational surveys and intelligence tests assess the higher-order cognitive abilities that are important for individuals to adapt to a globalised world, and for countries' economic advance. The globalisation of assessment is believed to have revealed the need to re-examine current educational practices in many parts of the world since they do not face up to the challenges posed by globalisation. Copyright © 2007 John Wiley & Sons, Ltd.

The central finding of Rindermann's paper is that tests designed to assess students' school achievement actually measure, together with intelligence tests, higher-order cognitive skills. Furthermore, factor analyses of the results from international student assessment and intelligence test studies revealed a strong *g*-factor, suggesting that cognitive ability differences between nations are unidimensional.

A methodological explanation advanced was that the questions asked in certain international student assessment studies, especially in OECD's PISA-Studies, are not intended to measure knowledge acquired in school, and are, therefore, not geared towards testing mastery of national curricula, but are instead intended to measure the degree to which students are able to apply to real-life problems what they have learned during the years of compulsory education (Adams, 2003). However, as Prais (2003) has argued, the PISA questions are nearer to tests of 'common sense', that is, of IQ, than to tests of educational attainment.

What dictates the need to give a lesser weight to *knowledge* questions and more weight to *thinking* questions in student assessment tests? It will be argued that these educational surveys are conducted in the context of globalisation, which characterises our era, and which has an impact on what is actually being assessed by some such tests (Kellaghan & Greaney, 2001). Globalisation is a process whereby countries and regions become more integrated due to the rapid movement of people, goods and ideas (Coatsworth, 2004). Today's youth will have to adapt to an increasingly complex world, i.e. to a new economic reality, defined by international trade and capital mobility, to the rapidly developing information, communication and media technologies, which facilitate communication of people across the world, to ever growing immigration flows as well as to the cultural transformations and exchanges that result from the above (Suarez-Orozco & Qin-Hilliard, 2004).

To deal with these challenges and adapt to such a demanding environment, young people will need to develop higher-order cognitive and interpersonal skills (Gardner, 2004). Such cognitive skills, i.e. reasoning, abstract thinking, etc., are assessed by PISA questions as well as by intelligence test questions.

Countries also need to adjust to the challenges posed by the process of globalisation. Countries' economies, which are directly tied to the global economy, require that

DOI: 10.1002/per

populations develop higher-order cognitive skills, and, in that respect, education and training are considered to be of critical importance for economic advance (e.g. Bloom, 2004; Burnett & Patrinos, 1996; Green, 1997). It has been argued that international student assessment plays a major role in ensuring that the outcomes of education are those that the economy needs and, furthermore, that it defines a country's position in educational achievement relative to that of its economic competitors, assuming that performance on such measures has implications for economic performance (Kellaghan & Greaney, 2001). Lynn and Vanhanen's (2002) argument that the economic situation of a nation is related to the average IQ of its population is in agreement with this line of thought.

A number of environmental and genetic explanations were advanced by Rindermann to account for the finding that cognitive ability differences between nations were uni-dimensional. The main thread of his arguments is that performance in both student assessment and intelligence tests seems to depend on similar environmental and genetic factors.

Rindermann examined the possible role of genetic factors to account for the homogeneity of national ability differences, and reported that inter-individual differences in cognitive abilities, as measured either by student assessment or by intelligence tests, have a large genetic component. However, as Sternberg, Grigorenko, and Kidd (2005) have shown, an attribute's heritability does not address its modifiability. The environment has substantial power to make a difference on learning skills, thinking skills, motivation to learn and on school performance (Sternberg & Grigorenko, 1999).

Education could then be examined as an important determinant of these cognitive ability differences between nations. It will be argued that the same forces that lead to the need for globalisation of assessment, also lead to a need for globalisation of education. As was argued in the previous section, international student assessment tests do not measure the degree to which students have memorised facts related to different disciplines, but assess instead students' abilities to apply concepts, skills and understandings to authentic problems that arise in real world settings (Adams, 2003). These thinking skills are required by the increasingly complex reality populations have to face as a result of the globalisation of the world. However, are educational systems around the world preparing children and youth to engage globalisation's new challenges, opportunities and costs?

Gardner (2004) argues that ministries of education and schools change at a 'glacial pace' and cannot follow the rapid social, economic and cultural transformations that are taking place in societies. Rote learning of factual and definitional information through repetition, drilling and solution of preconfigured problem sets, outdated curricula, unmotivated teachers and infrastructure difficulties are among the many problems besetting educational systems especially in developing countries (Bloom, 2004). It could be argued that such problems observed in many countries' educational systems are actually reflected in the assessments that test for higher-order cognitive skills.

Schools have traditionally taught children and youth the knowledge, skills and personal qualities, that will allow them to deal effectively as adults with the social, economic and political realities of their culture (Greenfield & Suzuki, 1998). Today's youth around the world have to face a more homogeneous global culture. Therefore, it is for the benefit of both individuals and societies that schools nurture a 'global citizen' with a well-developed ethnic identity. This would lead to a certain homogenisation of curricula that would emphasise the development of such cognitive and interpersonal skills as the ability to think analytically and creatively, to tackle problems flexibly, to tolerate differences and to collaborate with individuals from different cultures (Gardner, 2004), while at the same

DOI: 10.1002/per

time allowing room for maintaining and nurturing local knowledge, language and culture (Bloom, 2004).

The globalisation of assessment has revealed, then, the need to re-examine current educational practices in many parts of the world. The question is whether countries and their educational systems are ready to respond to the challenges posed by the inescapable process of globalisation.

Do Recent Large-Scale Cross-National Student Assessment Studies Neglect General Intelligence *g* for Political Reasons?

HELMUTH NYBORG

KF Andersen Leadership Academy, Lausanne, Switzerland
helmuthnyborg@msn.com

Abstract

Rindermann's analysis of international student assessment studies re-confirms previous findings that all intelligence- knowledge- and achievement-scales basically measure psychometric g. Several recent large-scale international assessment studies nevertheless chose literacy over psychometric g as their dependent variable, in order to better promote the notion of considerable student educational malleability. This politically correct choice compromises the studies and educational policy. Copyright © 2007 John Wiley & Sons, Ltd.

Rindermann (this issue) documents that student ability scales correlate highly significantly with each other within and across nations, and that all the intelligence-, knowledge- and achievement-scales basically reflect psychometric *g* in his analysis of data on millions of students in many countries.

This does not surprise seasoned psychometrics. Back in 1904 Spearman noted the positive manifold in student responses across different school subjects, i.e. all item responses correlated positively. Jensen (1998) refined Spearman's early two-factor analyses and stated that a second (or third) order general intelligence factor *g* always appears in ability tests (unless the mathematical procedure explicitly forbids it). The aggregation of data in huge cross-national samples with large test batteries reduces bias and increases reliability.

With this in mind, it is puzzling that recent international large-scale student assessment studies systematically avoid references to *g* research. The huge PISA2003 (OECD, 2003) study is a prime example, with no discussion of what stable individual difference in *g* could mean in terms of differentiated scholastic achievement (Nyborg, 2005). Moreover, Rindermann (this issue) aggregated their own data and finds—like Spearman did a century before—that different scales in different student assessment studies and different test ability approaches measure essentially the same construct, a strong *g*-factor. Rindermann (2006) also found a strong *g*-factor in a German state level analysis, as do Lynn and Mikk (2007) in the international TIMSS2003 study, and Nyborg (2005) does in PISA2003. We

DOI: 10.1002/per

then have a paradox: The *g*-factor obviously is *the* dependent variable in all the international student assessments studies, but the studies will not admit it!

Rindermann (this issue) explains the paradox noting that international student assessment studies seem driven by tradition, by personal provenience (lack of respect for alternative approaches) and by disciplinary climate (critics of PISA are degraded publicly at the expense of their scientific reputation, and aggressive mood often dominates, see Meyerhöfer, 2006). Rindermann also notes that the studies are carried out in Cupertino with political institutions and private companies, and the names of constructs, research content, names of countries and even addresses of test companies (Princeton), are influenced by political and economic interests. Obviously, such connections may compromise scientific objectivity.

There may be an even more serious hidden agenda lurking behind the systematic omission of *g* research (Nyborg, 2005). The *g*-factor is operationally defined as an estimate of a latent variable with numerous biological and brain correlates (e.g. Jensen & Sinha, 1993). It shows very high stability over many years (e.g. Larsen, Hartmann, & Nyborg, 2007) predicts trainability better than other factors (Ree & Earles, 1990), has impressive predictive validity for education and occupation (e.g. Gottfredson, 2003) and is highly heritable (e.g. Plomin & Spinath, 2004). Each of these characteristics would make the *g-factor* unsuitable as dependent variable in any projects designed to further student learning malleability without *a priori* theoretical or practical restrictions. In other words, it seems that the studies deliberately skip *g* because it would compromise their hope for fast changes in student and national performance.

I will be specific about this serious allegation. PISA20003 (OECD, 2003) has two main purposes: (1) to monitor how well national educational systems prepare their students and country for high productivity, and (2) to offer recommendations on how to improve individuals, schools and countries lagging behind. Obviously, such a study cannot operate with a dependent measure that is resistant to intervention at the individual, school or national level. To escape this trap, it introduces the concept of 'literacy' as its dependent variable. Literacy is '. . . concerned with the *capacity* of students to apply knowledge and skills and to analyse, reason and communicate effectively as they pose, solve and interpret problems in a variety of situations'. Literacy is '. . . much broader than the historical notion of the *ability* to read and write. It is measured on a continuum, *not as something that an individual either does or does not have*' but as '. . . a range of competencies . . . ' . . . '. . . rather than limiting the assessment to the possession of subject-specific knowledge'. (Emphasis added). PISA2003 further states that literacy is trainable, as 'The acquisition of literacy is a lifelong process—taking place not just at school or through formal learning, but also through interactions with peers, colleagues and wider communities'. Literacy is '. . . an indication of the learning development that has occurred since birth', and students' average performance at the national level '. . . depend on the quality of care and stimulation provided to children during infancy and the pre-school years, and on the opportunities children have to learn both in school and at home during the elementary and secondary school years'.

However, take away the largely unsupported malleability claims, and the definition of literacy is almost identical to Spearman's (1927) *g*-loaded 'Education of relations and correlates', such as '. . . reasoning to solve novel problems, as contrasted with recalling previously acquired knowledge or using already well-learned skills' (Jensen, 1998, p. 35). Tests that best reflect this require '. . . inductive and deductive reasoning, grasping relationships, inferring rules, generalising, seeing the similarity in things

DOI: 10.1002/per

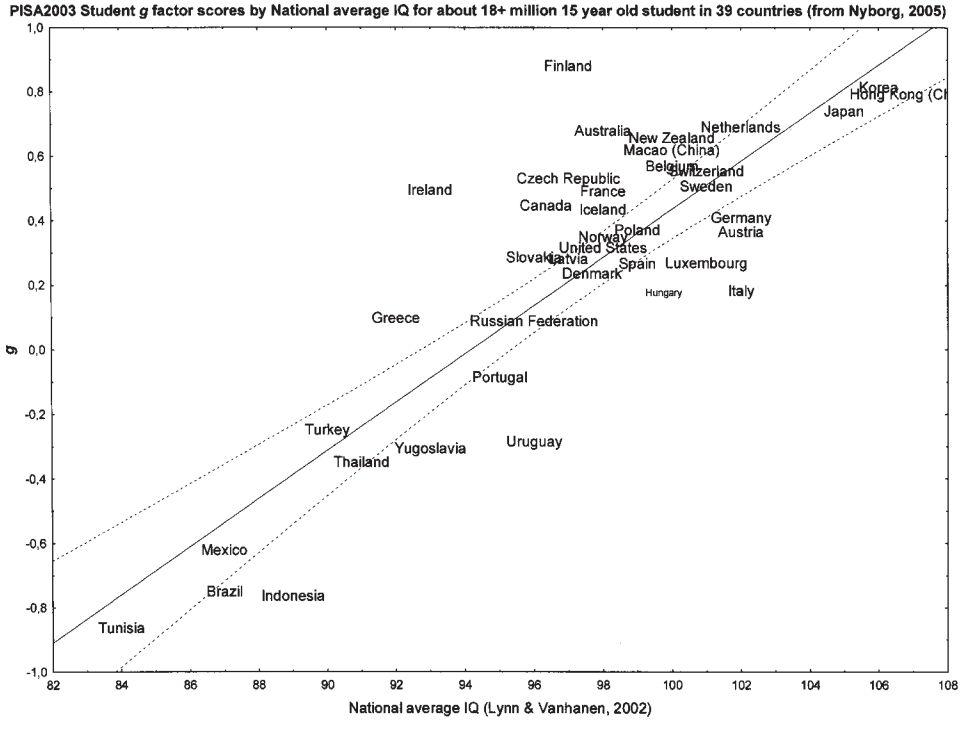

Figure 3. Cross-national correlation between IQ and PISA2003 *g* (adapted from Nyborg, 2005).

that differ . . . or the difference between things that are similar, . . . problem solving, decontextualizing a problem . . . '.

The fact that literacy equals *g* has two far-reaching consequences: The purpose of the assessment studies must be reformulated, and so must be the recommendations for changes. Figure 3 illustrates the point.

The figure plots national IQs (Lynn & Vanhanen, 2002) against student *g* (Principal unrotated Component) in PISA2003. Obviously, PISA2003 does not show how well national educational systems prepare their students and country for high productivity. Rather, the heavily *g*-loaded national IQ differences largely explain cross-national differences in the heavily *g*-loaded student assessments. Rindermann (this issue) speculates that heritability is very important in PISA2003 '. . . because of the strong *g*-factor and the relatively small impact of knowledge'. This restricts options for providing successful recommendations on how to make individuals, schools and countries lagging behind equal—except, perhaps, in terms of replacing low *g* populations with high *g* populations. To say otherwise comes close to committing academic fraud (Gottfredson, 1997a, 2000; Nyborg, 2003).

DOI: 10.1002/per

The Reciprocal Causation of Intelligence and Culture: A Commentary Based on a Piagetian Perspective

GEORG W. OESTERDIEKHOFF

Institute for Sociology, University of Karlsruhe, Germany
Oesterdiek@aol.com

Abstract

Social and cultural factors such as special child rearing practices and a lack of formalised education account for the stop of ontogenetic development below the level of formal operations in pre-modern or underprivileged social milieus. Only the socialisation practices in modern societies have been efficient enough to cause the IQ gains (Flynn effect) respective the growth of formal operations. Copyright © 2007 John Wiley & Sons, Ltd.

Rindermann (this issue) emphasises in his paper that different cognitive-ability approaches have been pointing to the same direction and have been coming to the same conclusions referring the relations of culture and cognitive ability. The different approaches widely confirm in identifying the relevant social and educational factors causing different ability levels, and some of these theories, as Rindermann especially emphasises, point to the social, economical, cultural and behavioural manifestations and consequences of different ability levels.

I would like to support these results by putting them into a broader and deeper perspective. First of all, there is to ask and enquire into the nature of cognitive ability. Is it an easily learned cognitive technique separated from the entire personality? Is it something like the ability to learn to drive a car or to play a flute? I maintain that this question can only successfully be answered within the frame of developmental psychology. Higher IQ scores and formal operations on the one side and lower IQ scores and pre-formal types of reasoning on the other side highly correlate, and both cognitive characteristics are stemming from the same source of cognitive and psychic maturation. According to the results of Piagetian stage theory, the sequence of sensory–motor, pre-operational, concrete operational and formal operational during childhood and adolescence unfolds not only the development of human cognition but the entire maturation of psyche and personality from suckling to adult (Piaget, 1950; Piaget & Inhelder, 1969). It is the entire human personality that is developing, not only some isolated cognitive techniques.

The low-IQ scores and pre-formal types of reasoning are therefore a result of lower levels of psychic-cognitive development, and the rise of intelligence reflects ontogenetic maturation. That is the reason why it is not so easy to raise somebody's intellect, let's say by some months or even a few years of intellectual training. A person must be educated all his or her youth by parental care, school attendance, social and cultural incentives and compulsions in order to attain high cognitive abilities. It is impossible to explain the development of these deeply rooted psychic processes without implying the role of the ontogenetic development of the nervous system exposed to cultural incentives and compulsions (Oesterdiekhoff, 1997, 2006a, 2006b).

DOI: 10.1002/per

Against this background it is easier to understand the nature of low intelligence of humans in underprivileged milieus, may be in poor milieus in developed countries or in backward milieus and regions in underdeveloped countries. The low scores respective the pre-formal types of reasoning in backward milieus are not due to biological factors but are stemming from the lack of sufficient parental instructions, school education, professional requirements and further cultural incentives. Adults in pre-modern societies collect their life experience within these more simple structures. They differ from children only due to their life experience, not due to qualitative structures of cognition. This fact is astonishing because we normally would assume that life experience alone might force qualitative changes. But especially the socialisation process in backward regions, the lack of stimuli (especially school attendance) in child development, usually prevents adults from reaching higher cognitive levels within their later life time (Dasen, 1977; Dasen & Berry, 1974; Hallpike, 1978; Luria, 1982; Oesterdiekhoff, 1997, 2006a, 2006b).

Rindermann mentions the social, health, economical, everyday life effects of different intellectual levels, both on the micro- and macro-sociological level. He is right in doing so, but again, to show this in the frame of Piagetian cross-cultural psychology reveals more insights and fruitful results. Pre-formal types of reasoning account for the magical-animistic world view of all pre-modern societies, for the belief in sorcerers and witches, archaic cults, worship of the dead and a lot of other forms of superstition. Pre-formal types of reasoning account for a lot of customs, manners and ideologies relevant to law, religion, economy, warfare, education and child rearing (Hallpike 1978; Oesterdiekhoff, 1997, 2006a, 2006b). The rise of modern societies is combined both with the rise of formal operations and IQ gains according to the well-known Flynn effect. Formal operations account for the development of sciences, enlightenment, modern law systems, economic growth, humanism, secularisation and a lot of other related processes. Formal operations among modern populations have caused the decline of magic and superstition that are so deeply rooted in pre-modern societies.

The relations between culture and cognition are so basing on reciprocal causations. The accumulation of knowledge, instructions, institutions, procedures in the sequence of generations surmounts a stage—especially in modern times—that makes it possible to socialise children that develop formal operations. And this cognitive development again promotes economic growth, professional skills, educational levels, social, cultural and moral developments of several kinds (Oesterdiekhoff, 1997, 2006a, 2006b).

How Smart are Nations? About Corrections and Correlations of International Data

MANFRED PRENZEL

Leibniz Institute for Science Education, University of Kiel, Germany
prenzel@ipn.uni-kiel.de

Abstract

The homogeneity of the results of international cognitive assessments which Rindermann claims in his paper will be questioned on two levels: First, Rindermann corrects national

Eur. J. Pers. **21**: 707–765 (2007)
DOI: 10.1002/per

data sets by a certain amount if specific test participation rates are available, without testing whether the samples are representative. Second, we will demonstrate with an example that the correlations reported can by no means be considered as evidence for the homogeneity of different cognitive tests and other indicators. Copyright © 2007 John Wiley & Sons, Ltd.

In his target paper, Rindermann would like to support the hypothesis that, on a macro-social level, different cognitive test instruments measure a common cognitive ability. For his analyses, he aggregates variables from large-scale assessments on the level of countries.

Rindermann's comments cause a feeling of déjà vu for me which is confirmed after close reading: To a large extent, the paper repeats arguments, data and analyses which have already been published in another paper (Rindermann, 2006). As there were two elaborate responses to this paper (Baumert et al., 2007; Prenzel et al., 2007), I do not wish to repeat the numerous objections argued in those responses. However, I do wish to comment on some information which was not given in the earlier publication (Rindermann, 2006). This concerns the corrections to original data carried out by Rindermann. Furthermore, I would like to point out how artificial *g*-factors can lead to correlations being misunderstood.

Practicalities: Can it be a bit more? International performance comparisons such as TIMSS and PISA are characterised by detailed technical reports. These present the extent to which the target samples were realised in the participating countries. In PISA, for example, the population is defined as 15-year-olds attending school. There are actually a few countries in which a large proportion of the population of 15-year-olds no longer attend school. These must be distinguished from the response rate of the samples which is on the basis of test participation. In accordance with the international guidelines, certain quotas (a total response rate of 85% of the target sample) must be achieved on the school and individual level, in order for a country's results to be reported. As these response rates vary between countries, Rindermann saw the need to correct the PISA data: for a response rate of 90%, he subtracted 20 points from the national average, with a rate of 80%, 40 points.

The fact that the Technical Advisory Groups of the international consortia have not yet even considered this procedure has its reasons. Different response rates do not mean right from the start that the samples are distorted. There are numerous reasons why schools or individuals do not participate in the test. There are also possibilities for testing the extent to which the samples are representative and these possibilities are (as is the case in PISA) consistently applied (Adams, 2003). We thus used additional information from the sampling frame (the school grades of the students) in order to test sample distortions due to differential test participation in Germany. In PISA 2003, test participation varied between the 16 federal states from 85 to 96%. Our analyses did not provide any indication that the sample was distorted (Prenzel, Drechsel, & Carstensen, 2005); nevertheless, scores of several states would have been heavily punished through Rindermann's procedure. With his corrections, Rindermann accuses samples in PISA of not being representative without having actually tested this.

Similarly, Rindermann corrects national data from IEA studies if the average age was above that of the international sample criterion. By referring to estimates (on the basis of

DOI: 10.1002/per

cross-sectional data) of competency increases in the course of a school year, Rindermann corrects the data substantially ($d = 0.42$). However, in longitudinal studies, much smaller performance increases can be seen per school year (Prenzel et al., 2006). In contrast, Rindermann does not bother with any corrections to the IQ data, probably because in the studies summarised by Lynn and Vanhanen (2002, 2006) no comparably detailed reports on response rates are available. Rindermann's corrections are therefore not carried out consistently. Imagine if the OECD or the IEA, criticised by Rindermann, were to round country results up or down according to average age or test participation as Rindermann does!

Conceptual factors: What correlations and general factors can mean. Correlations are a helpful tool for describing coherences in data sets. Although it has long been common knowledge that correlations cannot be interpreted causally or substantially, they do seem to influence assumptions about what was measured and what could account for the results.

In order to show that wrong conclusions can easily be made from correlations between variables of cognitive assessments, I would like to outline an example: Somebody could come up with the idea to correlate the technical variables of cars. For example, the cubic capacity and the amount of cylinders, the horse power, the CO_2-emission and the petrol usage. We could include numerous other variables into our data matrix which would also correlate highly with these data. We could easily identify a *g*-factor here. However, if we consider the variables addressed, we are dealing with input variables (petrol usage), system variables (cylinders and cubic capacity) and output variables (CO_2-emission). Thus, what does the identification of a *g*-factor mean in this example? It reduces information but does not give any indication of a useful substantive entity behind these measures. In this example, the general factor stands for an incorrect concept. Incidentally, the correlations in this example may be very high on the 'individual level' (the cars), but in no way perfect. From the aspect of technical efficiency and impact on the environment, these small differences gain considerably in importance. This example is also suitable for imagining a correlation achieved by an aggregation of the variables at a country level: similarly, we can form national averages for cubic capacity, cylinders, CO_2, etc. However, the aggregation on the country level obscures the small important differences more than ever.

What does this example mean for the interpretation of data from various different cognitive tests? It has become obvious that, for a fairly long time, we have an international curriculum, i.e. all countries have a comparable mathematics, science and reading curriculum. Every year, 'Education at a glance' (OECD, 2006) shows how similar international timetables are. No matter where children grow up nowadays, they are confronted with this international curriculum, one could say, with a 'construction plan' of knowledge development. But, in keeping with our example, how should we understand the cognitive variables in Rindermann's model? What are the input, system and output variables in this case? Is the intelligence present in the system or does the system need input? Or do our tests only measure output? Do Rindermann's intelligence tests, termed as 'school near', measure the output and the 'school distant' tests the input? If this is the case, did TIMSS and PISA in fact test intelligence? (No! and Rindermann should have mentioned this.) So, how should we interpret the general factor determined by Heiner Rindermann? In exactly the same way as the 'everything involving a car' general factor!

Finally, it must be said that large-scale assessments contain interesting information about relevant competency differences if we analyse the data carefully, for example, by also considering differences in the average or by looking at the results of subgroups.

DOI: 10.1002/per

Generalisability, Groups, and Genetics

J. PHILIPPE RUSHTON

Department of Psychology, University of Western Ontario, Canada
Rushton@uwo.ca

Abstract

Rindermann shows that g is highly generalisable. We can add: (a) predictive validities generalise across cultures; (b) g-loaded items found relatively difficult by the Roma (Gypsies) in Serbia are found relatively difficult by East Asians, Whites, South Asians, Coloreds and Blacks in South Africa and (c) group differences are more pronounced on more heritable items, indicating they are partly genetic. Copyright © 2007 John Wiley & Sons, Ltd.

Rindermann (this issue) has provided a compelling integration of data. His results join those from industrial/organisational psychology, cross-cultural psychology, evolutionary psychology and behavioural genetics to show that GMA is a human universal. As he rightly says, there is no need for conceptual differences among so many sub-disciplines.

The evidence for generalisability is even stronger than Rindermann describes. For example, predictive validity is high across diverse cultural groups. Sternberg et al. (2001) found that GMA in Kenyan 12- to 15-year-olds predicted school grades at the same levels they do in the West (mean $r = .40$). Rushton, Skuy, and Bons (2004) found that GMA predicted university performance equally well in African and non-African engineering students ($r \sim .30$). Salgado et al. (2003) demonstrated the international generalisability of GMA across 10 member countries of the European Community, thus contradicting the view that criterion-related validity is moderated by differences in a nation's culture, religion, language, socio-economic level or employment legislation. They found scores predicted job performance ratings .62 and training success .54. The validities were the same, or even higher, than those reported in the US, where there is again a quite different corporate culture, mix of populations and legislative history.

As the trend towards a more global economy continues, population differences in mean GMA are likely to become more salient, both within and across countries. To examine the validity of Lynn's (2006; Lynn & Vanhanen, 2006) IQ map of the world (see also Rindermann's Figure 2c), I travelled to Serbia and South Africa to meet new colleagues and collect more data. In South Africa, we tested several hundred Black university students and found an average IQ of 85 (Rushton et al., 2004). This confirmed Lynn's estimate of an average IQ of 70 for sub-Saharan Africa when it is assumed that Black undergraduates average 15 points higher than the general population, as their counterparts do in the West. In Serbia, we tested several hundred adult Roma (Gypsies), a diverse population of South Asian stock, and found an average IQ of 70 (Rushton, Cvorovic, & Bons, 2007). This confirmed Lynn's estimate of an average IQ of below 90 for the South Asian population, although our scores were much lower than expected. The data also showed the African/non-African and Roma/non-Roma differences were more pronounced on *g*; there was no evidence of any idiosyncratic cultural effect.

Group differences in GMA are also heritable. Rushton and Jensen (2005) examined 10 categories of technical research to conclude that in the US, East Asian–White–Black IQ

DOI: 10.1002/per

differences were from 50 to 80% heritable, just as individual differences are within a group (Bouchard & McGue, 2003; Jensen, 1998). The evidence included: (1) the IQ distribution around the world is consistent across time and place; (2) the race-IQ difference is more pronounced on the more *g*-loaded subtests; (3) the race-IQ difference is more pronounced on the more heritable subtests; (4) the race-IQ difference is paralleled by brain size differences; as well corroborating studies of (5) racial admixture; (6) trans-racial adoption; (7) regression to the mean; (8) 60 related *r-K* life-history traits; (9) human origins research and (10) the inadequacy of culture-only explanations (see also Rushton, 2005).

Most recently, Rushton, Bons, Vernon, and Cvorovic (2007) examined two independent twin studies to further test the hypothesis that genes influence group differences in about the same proportion as they do individual differences within a group (i.e. about 50%). We estimated the heritability of scores on the diagrammatic puzzles of the Raven's Progressive Matrices, a well-known, culture-reduced test of GMA. In Study 1, the heritabilities were calculated from 199 pairs of 5- to 7-year-old monozygotic (MZ) and dizygotic (DZ) twins reared together from the Western Ontario Twin Project. In Study 2, the heritabilities were calculated from 152 pairs of adult MZ and DZ twins reared apart from the Minnesota Study of Twins Reared Apart. In both studies, the group differences were more pronounced on the more heritable items. In Study 1, the comparison was between the 5- to 7-year-old twins and 94 adult Roma in Serbia ($r = .32$; $N = 36$, $p < .05$). In Study 2, there were 11 diverse groups: the twins reared apart; another sample of Serbian Roma and East Asian, White, South Asian, Coloured and Black high school and university students in South Africa. In 55 comparisons, the heritabilities correlated with the magnitude of the group differences on the same items (mean $r = .40$; Ns $= 58$, ps $< .05$), indicating the differences are partly genetic.

In conclusion, the results show that both individual and group differences are part of the normal variation to be expected within a universal human cognition, located on *g*, and caused by genetic as well as environmental influences.

Profiting From Controversy

MANFRED SCHMITT

Department of Psychology, University of Koblenz-Landau, Germany
schmittm@uni-landau.de

Abstract

My comment will address the scientific value of Rindermann's contribution, wrong conclusions that might be drawn from it, and his quest for interdisciplinary co-operation.

Scientific value. Rindermann's analysis is valuable for several reasons. He offers a large-scale cross-national analysis of cognitive achievement data. Altogether, this analysis includes more countries, constructs, age groups, grade levels, assessment paradigms and participants than has any previous analysis. Such a comprehensive review is, in itself, a valuable contribution to the literature.

DOI: 10.1002/per

Rindermann reflects carefully upon several biases that endanger the internal validity of his analyses: selective participation, systematic differences between countries in age and grade level, school attendance and participation rates. Even though Rindermann's manner of handling these biases may not convince all readers, addressing them is important in order to sensitise readers who are either not aware of such biases or who do not have the methodological expertise to understand the effects they might have.

Rindermann implements correctional procedures for some biases. These corrections are crude but transparent and thus challengeable. Severe correction mistakes are unlikely because the analyses of corrected and uncorrected data result in rather similar correlational patterns. Uncorrected and corrected correlations differ substantially in only a few cases. These cases could be analysed in more depth using additional analyses and additional data.

Rindermann's conclusions are provocative and will stir controversy. This seems likely given reactions (Baumert et al., 2007; Prenzel et al., 2007) to another recent analysis of student assessment data at the level of the individual by the same author (Rindermann, 2006). In line with the main message of his present paper, Rindermann (2006) suggests that instruments used in PISA and TIMSS share a single common factor and that this factor is the same factor that intelligence tests share—whatever it is and however we want to name it (e.g. intelligence, cognitive ability, literacy and intelligence plus knowledge). Controversy is good for scientific progress! Other controversies such as the person versus situation debate have had very profitable outcomes for our discipline (Kenrick & Funder, 1988).

Controversies contribute to scientific progress for several reasons. For example, they force opponents to sharpen their arguments and they stimulate additional research. In combination, both of these effects result in more sophisticated theoretical models, better research designs, more appropriate data analyses and more careful conclusions. Eventually, controversies may generate sufficient data for quantitative meta-analyses that are less biased than single studies. The person versus situation debate eventually lead to a meta-analysis by Richard, Bond, and Stokes-Zoota (2003) that included more than 16 000 independent data sets.

Risks. Readers may conclude that the factors responsible for the convergence of cognitive achievement measures at the national level will be the same as those that are responsible for convergence at the level of individual performance data. This conclusion is unsafe. It may be true, partially true but also wrong. The two sources of variance, national and individual, are logically independent of each other. Let me illustrate the point with a simple hypothetical example that begins with five assumptions: (a) all tests considered share the same two common sources of variance, genetic differences and environmental differences; (b) both factors (genes, environment) have equal effects on performance; (c) both factors are independent; (d) environmental conditions vary only between nations and (e) genetic outfit varies only between individuals within nations. If these assumptions were met, a similar degree of convergence among the tests would appear at the national level and at the individual level. The causes of convergence, however, would be entirely different. Convergence at the national level would be caused exclusively by environmental differences. Convergence on the individual level would be caused exclusively by genetic differences. Although this example is made up and unrealistic, the message is important: similar or even identical patterns of convergence can have entirely different causes.

Wrong conclusions may also result from a failure to understand the logic of aggregation. Aggregation always implies a loss of information. It contributes to the robustness of data patterns. This effect is desirable in some cases but undesirable in others, simply because

DOI: 10.1002/per

the cost of robustness is a lack of sensitivity. Aggregation is a good strategy if measures are sensitive to random error or irrelevant sources of variance. Aggregation is a bad strategy if measures are sensitive to systematic factors that have the potential to reveal the causal mechanism that generated the data. I am not saying that Rindermann aggregated wrongly. I am saying that every aggregation requires very careful theoretical consideration of the possible meaning and explanatory value of the information that is lost by aggregation.

Quest for Cupertino. I agree with Rindermann's quest for interdisciplinary Cupertino. Each discipline has a limited set of theories and methods for understanding the phenomena of interest. Therefore, each discipline can add incremental validity to a comprehensive theoretical model of that phenomenon. The value of critical multiplism (Cook, 1985) is not limited to multi-method assessment. It includes theories, research questions, research strategies, sources of information, and strategies of integrating and condensing information.

The Softer the Truth, the Harsher the Fight

BIRGIT SPINATH

Department of Psychology, University of Heidelberg, Germany
birgit.spinath@psychologie.uni-heidelberg.de

Abstract

Rindermann's target paper is bound to provoke retorts. The present commentary analyses the explosive potential of Rindermann's thesis along five questions: Are international student assessment tests flawed? Should we give up the distinction between intellectual potential and scholastic performance? Is intelligence heritable but school achievement is not? Is school achievement primarily dependent on the educational environment but intelligence is not? Can we advance something that is heritable? Copyright © 2007 John Wiley & Sons, Ltd.

Rindermann closes his target paper complaining about an aggressive mood in the discussion about student assessment studies. Indeed, his recent papers instigated a heated argument. But what exactly does provoke so many objections and why does the debate even abandon a pure matter-of-fact line of reasoning? Recipients might react to two aspects: First, one might disagree with the methods used to reach the main conclusions and, second, one might dislike the implications that followed if Rindermann's thesis was true. In my view, the reasons for the sharpness of the argument have not so much to do with the adequateness of the methods used by Rindermann but with the implications of his thesis.

Rindermann's thesis and its implications. The main thesis of Rindermann (this issue) is that 'student achievement assessments and intelligence tests primarily measure a common cognitive ability *at the macro-social level*'. This thesis adds to the message of a recently published paper by Rindermann (2006) in which he stated that, *on an individual level*, international student assessment studies measure a *g*-factor of cognitive abilities. In the following, the implications of this thesis are analysed along five questions.

DOI: 10.1002/per

Is there a fundamental flaw in the design and interpretation of international student assessment tests? Surely for those researchers who planned, conducted and reported on international student assessment studies, Rindermann's work implies the accusation of not doing a good job. This alone would be sufficient to explain sharp reactions. But whether the data provided by Rindermann prove those responsible for student assessment studies wrong depend on whether the seemingly contradictory interpretations of the data are conciliable. Rindermann provides strong evidence for one general factor on which most of the school achievement tests and intelligence tests load very highly. Nevertheless, the existence of a general factor does not conflict with the fact that more specific factors may explain even more variance in school achievement tests and that multifaceted theoretical models are adequate to describe the data (Baumert et al., 2007; Prenzel et al., 2007).

More important than the adequate number of factors to describe the data seem to be the question of what exactly these factors measure. According to Rindermann, the general factor can clearly be interpreted as general intelligence, whereas international student assessments claim to measure competencies acquired in school. Whether these views are reconcilable depends on the definitions of intelligence and competencies. If intelligence is defined as the ability to think, this definition does embrace all cognitive competencies, no matter how context-specific these competencies are. In this case, school achievement tests *have to* measure intelligence to be valid, because intelligence is the most important prerequisite to acquire school-based competencies. Nevertheless, there is a discomfort in collapsing intelligence and school-based competencies, which leads us to the second question.

Should we give up the distinction between intellectual potential and scholastic performance? If it was true that intelligence tests and school achievement tests basically measured the same latent construct, we had to give up the distinction between intellectual potential and scholastic performance. This would impose a problem not only for the construct and measurement of school achievement but even more so for the construct and measurement of intelligence. Intelligence tests claim to measure a potential underlying manifest achievement. They are employed to investigate whether the reason for poor school achievement is a lack of cognitive ability or something else. It is an unresolved paradox that, in order to develop an instrument that measures intellectual potential, this test has to be validated via strong but not too strong associations with actual school performance. The distinction between intellectual potential and scholastic performance is theoretically sound and builds an important basis for educational diagnosis and counselling. In his target paper, Rindermann suggests to simply rename what is measured by student assessment tests as 'common cognitive ability'. In the light of the aforementioned, I believe this does not solve problems but adds even greater ones.

Is intelligence heritable but school achievement is not? Is school achievement primarily dependent on the educational environment but intelligence is not? Can we advance something that is heritable? One might see the explosive potential of Rindermann's thesis in the fact that we have extensive evidence of the heritability of intelligence (Plomin & Spinath, 2004), meaning that a substantial proportion of the differences between individuals is attributable to differences in genes (for problems with transferring the logic of heritability to a national level see the commentary by Frank M. Spinath). The stream-of-thought would be something like this: 'If student assessments measured intelligence and intelligence is heritable, then the differences between individuals (and nations) are fixed and are not malleable by educational processes'. This stream-of-thoughts contains three misconceptions.

DOI: 10.1002/per

First, as Rindermann points out, by now we have evidence that whatever it is that school achievement tests measure, it is as heritable as intelligence (e.g. Bartels, Rietvelt, van Baal, & Boomsma, 2002). Second, the fact that differences in school performance are heritable does not mean they are not affected by school environments. On the contrary, behaviour genetic studies provide the best evidence for the importance of the environment in shaping individual differences, because they control for the effects of genes. Behaviour genetic studies show that the environment, although we do not know which environment, is as important for differences in school achievement as it is for intelligence. Without a good educational environment, no individual or nation can reach a high level of what is measured in school achievement tests. Third, heritability is only a bug-a-boo if it is equated with non-malleability. This common misconception roots in a lack of discrimination between mean-level changes and correlations. Heritability coefficients tell that even in perfect educational environments, there would still be differences between individuals. But they do not tell us to what extent the mean levels of all or one single individual are malleable.

Conclusion. Many of the implications of Rindermann's thesis are less far-reaching than might be assumed. Others have even more severe implications than Rindermann might have intended. Empirical data and logic argumentations should be the best weapon in the continued quest for truth.

International *g*: Mixed Feelings and Mixed Messages

FRANK M. SPINATH

Department of Psychology, Saarland University, Germany
f.spinath@mx.uni-saarland.de

Abstract

This commentary reflects on the misleading amalgamation of behaviour genetics findings on the aetiology of individual differences in intelligence and speculations about the causes of international mean differences in intelligence at the group level. It questions the scientific value of mapping or ranking national mean IQs given that the vast majority of the variation lies on the individual level. Copyright © 2007 John Wiley & Sons, Ltd.

I have been following the recent publications of Rindermann on (international) cognitive-ability comparisons with great interest. The present paper as well as an earlier contribution (Rindermann, 2006) are intriguing and thought-provoking, and the data presented appear to be solid although I do not pretend to have read the originals of more than a handful of the papers or books cited. I also find it difficult to judge the adequacy of the presented corrections to increase the comparability among the various international student assessment and intelligence test studies. But neither do I doubt the soundness of the structural analyses or the validity of the national IQ data borrowed from Lynn and Vanhanen (2002, 2006) in general nor do I object to the idea that a positive manifold exists between student assessments and intelligence tests.

DOI: 10.1002/per

Results from research in the field of behaviour genetics and intelligence are largely compatible with the notion of a broader *g-factor* that includes not only psychometric intelligence in the narrow sense but also psychophysical, educational as well as performance and achievement indicators, and even motivational measures. Furthermore, multivariate genetic research indicates that genetic *g* (that is, pleiotropic gene effects) accounts for a large portion of the genetic variance of diverse psychometric cognitive tests (e.g. Plomin & Spinath, 2002).

I do, however, find it necessary to comment on the misleading amalgamation of findings from genetic studies at the individual level and speculations about the causes of international mean differences at the group level. Despite the fact that Rindermann makes such a distinction when he discusses genetic components, the (fundamental) difference between these issues is not properly reflected throughout the paper (e.g. 'Causes of the high correlations are seen in (. . .) common developmental factors at the individual and national levels including known environmental and unknown genetic influences'. A critical comment seems especially warranted because most authors known for their empirical research on international mean differences in IQ (e.g. Lynn, Rushton, or Templer) are also known for their interest in race differences in intelligence (e.g. Lynn, 2006), no doubt a delicate issue.

It therefore needs to be emphasised that the vast majority of genetic variation is not between populations, but among individuals within them. If we are interested in the variation in the intelligence of humans and its aetiology, we need to focus on the variation *within* populations (e.g. Loehlin, 2007). Virtually all behaviour genetic studies including the ones cited in Rindermann's paper do not address mean differences at all and there is no direct link between the concept of heritability and mean-level differences. Univariate behaviour genetic analyses (e.g. in twin studies) typically decompose the observed variance into a genetic component (heritability or h^2) and two environmental components (shared (c^2) and non-shared (e^2) environment). If twin studies incorporate repeated measurements, changes in individual scores as well as group level changes from Time 1 to Time 2 may occur. However, neither do they depend on nor can they be derived from the relative importance of genes and environment at either measurement occasion.

Multivariate analyses can provide additional information using cross-twin cross-time (or cross-trait) data. In a nutshell, genetic influences are suggested if Twin A's intelligence score at Time 2 can be predicted from Twin B's intelligence score at Time 1 and if this cross-correlation is higher in MZ compared to DZ twins. Thus, multivariate analyses can provide information on genetic contributions to rank order stability (if applied in longitudinal settings) or to the covariation of traits, if different constructs are measured. Even multivariate twin analyses, however, do *not* provide an explanation for mean-level differences between measurements or mean-level changes that might occur over time.

Twin data are occasionally used for mean-level analyses when genetic contributions at the extremes are of interest (e.g. in research on mild mental retardation; see Spinath, Harlaar, Ronald, & Plomin, 2004). However, the particular methodological approach used here (called DeFries–Fulker extremes analysis; DeFries & Fulker, 1985), is not applicable to the comparison of mean-level differences between distinct groups.

In short, the well-documented and highly consistent result that genes contribute to individual differences in intelligence can not be used as evidence for genetic influence on group differences in intelligence.

Nevertheless, Lynn's (2006) book on race differences in intelligence closes with the confident assertion that differences between countries and their indigenous populations are

DOI: 10.1002/per

genetic. Although Rindermann's own treatise of the potential causes for different cognitive-ability levels of nations is more cautious (he lists genetic causes among others in his suggestions for future research), the overall message of his paper(s) remains mixed. What is the scientific value of world maps of IQ differences or figures displaying negative correlations of national cognitive level and, for example, percentage of muslims in a country (Rindermann, 2006, p. 83)?

Correlations between mean differences in intelligence (be it on the regional level or on the level of nations) and macro-social attributes hardly elucidate processes and psychological mechanisms involved and, again, ignore the vast variation within the observed groups. Furthermore, decades of behaviour genetic research on intelligence indicate that from early adulthood onwards environmental influences operate primarily in a non-shared fashion, that is, contribute to differences rather than similarities of individuals reared together. Interdisciplinary efforts are required to study such influences in genetically informative designs. It remains to be seen whether international comparisons and macro-social correlations prove to be helpful data in this context or rather turn out to be ideological baggage which the field can much better do without.

Abilities as Achievements, or is it Achievements as Abilities, or is it Both, or Neither?

ROBERT J. STERNBERG

Tufts University, USA
robert.sternberg@tufts.edu

Abstract

Rindermann's paper shows that well-known ability and achievement tests all roughly measure the same thing, general ability. Three potential implications are that the distinction between ability and achievement is not clear, that we should use broader psychological theories on which to build tests, and that we should consider teaching for leadership rather than merely for academic facts and skills. Copyright © 2007 John Wiley & Sons, Ltd.

Rindermann (this issue) has provided a comprehensive and compelling review of the literature on the relationships of various tests of abilities and achievement, including such high-profile assessments as PISA and TIMSS. There are many findings, but three stand out:

1. All the tests of abilities and achievements reviewed are highly correlated, whether or not they are supposed to measure the same thing.
2. All the tests of abilities and achievements reviewed are highly *g*-loaded.
3. Although these tests form a single factor, wealth and democracy form separate factors.

The question then arises of what we are supposed to make of all this, and even of whether these results are good news or bad news. Those who publish and use these tests

DOI: 10.1002/per

perhaps should feel some discomfort with these results, because the results suggest it does not matter a great deal what test you use—whether ability or achievement—or even what sort of achievement you are going to measure. It is possible to create ability and achievement tests that measure more distinctive qualities, such as creative and practical skills (Hedlund, Wilt, Nebel, Ashford, & Sternberg, 2006; Sternberg & the Rainbow Project Collaborators, 2006; Stemler, Grigorenko, Jarvin, & Sternberg, 2006), but until we do so, we are left with tests in widespread use that, in some respects, are rather narrow in their focus. As things stand, all of the ability and achievement tests currently in widespread use measure more or less the same thing, even though basing tests on a broader theory of abilities and achievement might create differentiation (e.g. Gardner, 1983; Sternberg, 1997). What do we make of the current situation?

One interpretation, supported by the data, is that tests of abilities are really tests of achievement—that, in essence, they are measures of developing expertise (Sternberg, 1998a, 1998b, 1999a, 1999b, 2001). Almost certainly, this is true to some extent. Ability tests often are achievement tests for achievements that one was supposed to have attained some years before. They include assessments of vocabulary, arithmetic skills, comprehension and other skills that, directly or indirectly, are taught in school. The content is largely the same as that on achievement tests. Indeed, one well-known test used widely for university admissions in the United States—the SAT—has been called, at various times, Scholastic Ability Test, Scholastic Assessment Test, SAT Reasoning Test and just plain SAT. Some have referred to it as 'Scholastic Achievement Test', although that is not an official name. The SAT-I is the reasoning portion and the SAT-II the achievement portion, although many items are practically the same. So, on this interpretation of the Rindermann results, ability tests are merely achievement tests, if, perhaps, in a (very) thinly disguised form.

A second interpretation is that tests of achievement are really tests of abilities. This interpretation is also correct, in the sense that ability test scores predict achievement test scores about as well as achievement test scores. The Rindermann results suggest it does not much matter what you call the test—they all measure about the same thing. Indeed, Frey and Detterman (2004) found that IQ tests and SATs measure roughly the same thing, despite the care that the College Board, creator of the SAT, has taken to emphasise that the tests are not the same.

On either of these two interpretations, it is not clear why we would need separate tests of abilities and achievement. At best, achievement tests measure a more advanced level of academic expertise than ability tests, but all measure roughly the same kind of expertise.

A third interpretation is that both ability and achievement tests measure roughly the same thing, but both are dependent on some third factor or set of factors. This interpretation is also correct. Without doubt, both tests are dependent, for example, on amounts of schooling and the achievements that result from such schooling. Indeed, Rindermann cites research showing that each year of schooling adds a substantial amount to IQ, with the amount depending on the particular study (e.g. Ceci, 1991; Winship & Korenman, 1997). So schooling not only increases achievement, but also ability conceptualised, like achievement, as developing expertise.

An important finding of the Rindermann analysis is that democracy and wealth are factors that are largely distinct from the factor of general cognitive ability. Our societies tend to place a great deal of emphasis on cognitive skills. But given that abilities are not terribly predictive of wealth or democracy, perhaps our focus ought to lie elsewhere—in producing wealth and democracy, not just higher levels of *g*. I have suggested that the

DOI: 10.1002/per

focus of education should be on leadership—on creative skills and attitudes to come up with new ideas, analytical skills and attitudes to ensure that they are good ideas, practical skills and attitudes to implement the ideas, and wisdom in order to ensure that the ideas contribute to the attainment of a common good (Sternberg, 2005, 2007).

Methodological and Conceptual Notes on National Cognitive Ability

THOMAS VOLKEN

Sociological Institute, University of Zürich, Switzerland
volken@soziologie.unizh.ch

Abstract

While at first sight, the concept of national cognitive ability is appealing, a closer inspection reveals major conceptual and methodological problems which need to be addressed. In particular, the meaning as well as the scope and reach of the concept remain vague, the aggregation process does not consider estimates of different precision, and (too) many values are estimated. Copyright © 2007 John Wiley & Sons, Ltd.

Cross-national comparison of cognitive ability is a complex and demanding task. Not surprisingly, this research field has been largely neglected. More specifically, very few studies exist which take the state (country) as their unit of analysis, and those which have done so have attracted considerable critique (e.g. Lynn & Vanhanen 2002, 2006).

Rindermann takes this research strand further in two important ways. First, different measures of cognitive ability are incorporated in order to demonstrate a *g*-factor of differences between countries. Second, a new country-level indicator for 194 countries—termed *common (national) cognitive ability*—is constructed from different measures of cognitive ability.

While choosing and constructing his indicators, Rindermann very carefully addresses many important conceptual, methodological and design issues which usually give researchers interested in a comparative cross-national approach to cognitive ability a hard time. More specifically, the country mean scores of different student assessment tests are corrected for age and school participation in order to enhance comparability.

Without doubt, Rindermann's analysis has many merits. However, two major problem areas—one methodological, the other conceptual in nature—need to be systematically tackled. I shall start with a brief review of the methodological problems.

First, the method of building country-level indicators consists of a stepwise aggregation process of *mean* test scores across various measures of cognitive ability in various points in time. Essentially, this process implies that all studies in all points of time in all countries can be assigned an equal weight. Yet, this assumption is hardly met. Most importantly, studies considerably vary in their sample size. By assigning equal weight to small and large sample studies, the resulting mean effect size may be substantially biased because different sample sizes produce estimates of different precision. One potential solution to

DOI: 10.1002/per

this problem would be to draw from the experience of meta-analysis and use the inverse variance weight as a mechanism for correcting effect sizes over studies, time and space in order to make combined estimates more precise (Hedges & Olkin, 1985; Lipsey & Wilson, 2001; Wolf, 1986). Higher weights would then be given to studies with larger samples (smaller sampling errors), lower weights to those with smaller samples.

Second, only one third (32.5%) of the mean scores calculated for common cognitive ability are based on student assessment and intelligence test studies. Roughly one third is based on either intelligence tests (25.8%) or student assessment test (7.7%) and the remaining part (34%) of all country scores was estimated. Essentially, this means that two thirds of all common cognitive country scores are either based on estimates which assume *ceteris paribus* conditions for various country characteristics *and* the research process itself or the scores are mainly based on intelligence test studies. The latter, however, are often problematic because of their small sample sizes as well as for the methodological reasons outlined by Rindermann. In sum, common cognitive country scores run a considerable risk of being biased.

Third, the scope and reach of the common cognitive-ability indicator is difficult to interpret. Because the underlying indicators which constitute the syndrome stem from different age cohorts and are reported for different years, it is unclear to what timeframe and population the indicator refers. While for some research contexts, this will not pose an essential problem, for others—mainly those involving the analysis of causal effects—this will be a major obstacle.

I shall now turn to the second problem area and discuss some conceptual issues which need clarification. Rindermann claims that different cognitive-ability studies measure at the macro-social level a common (national) cognitive ability. Furthermore, he claims that this ability consists of the ability to think (intelligence) and the ability to acquire, store and use relevant knowledge.

These claims imply strong conceptual assumptions. Most importantly, equivalent micro- and macro-level processes of cognitive-ability structuration are assumed. In essence, this means that the concept of cognitive ability is the same across levels. But is this really the case?

I doubt that the concept of cognitive ability is so easily transferable between micro and macro levels. On one hand, individuals are agents in their own right; macro-level entities—such as nations, states or societies—can only become agents through individuals. This means that macro-level entities cannot be intelligent, unless one is referring to intelligence metaphorically. On the other hand, macro-level entities are heavily involved in creating, storing and using knowledge. States, for example, are essential for creating, facilitating and regulating diverse knowledge infrastructures, including media (newspaper, television and internet), libraries, universities, think tanks and innovation clusters such as Silicon Valley. Although these social institutions and structures greatly influence individuals and are reciprocally influenced by individuals, macro-level knowledge structures are quite different from their micro-level counterpart. To name just two differences: the former typically have a much longer life span and their operations are heavily based on (bureaucratic) rationality.

Given the absence of an agent in the case of macro-level entities and the many qualitative differences in the way knowledge is stored, used and processed on the micro-level as compared to the macro-level, it would either be appropriate to drop the term cognitive ability on the macro level or to construct the concept in a way that productively links the two levels together. Personally, I am inclined to opt for the former because on one

DOI: 10.1002/per

hand, the 'unambiguous' micro concept of cognitive ability could be preserved. On the other hand, the methodological problems outlined above would not apply.

Of course, this implies that researchers would have to base their analysis on the corresponding micro data. While cross-national or other higher-level effects could still be studied by application of multi-level analysis, the decision to drop the macro-social concept of cognitive ability would severely limit the universe of future research questions outlined by Rindermann. Yet, I am very sympathetic to his idea of extending research on cognitive ability beyond the individual and beyond the narrow disciplinary boundaries.

Something Old, Something New, Something Borrowed and Something Blue: National IQ and the Integration of Cognitive-Ability Research

MARTIN VORACEK

Department of Basic Psychological Research, University of Vienna, Austria
martin.voracek@univie.ac.at

Abstract

This commentary focuses on methodological issues encountered in estimating national IQ; specifically, on criticism regarding one source of national IQ figures (Buj, 1981), which criticism appears unjustified; and affirms Rindermann's call for a broad integration of different paradigms and lines of cognitive-ability research, illustrated with an example from suicide research. Copyright © 2007 John Wiley & Sons, Ltd.

Rindermann's (this issue) profound analysis of national IQ and scores on student assessment studies, following up and extending Rindermann (2006) and showing high to very high positive aggregate-level associations between these, is an intriguing and truly original contribution. I opine this finding will have repercussions on various fields (intelligence research, differential and educational psychology, psychological assessment, economics, policy decision-making, migration research, human biology and ethology, evolutionary cognitive psychology, behavioural and population genetics) and will surely inspire a wealth of further inquiry along these lines. The author is to be congratulated. Anticipating specific commentaries from authorities of intelligence and educational research on the target paper, I limit myself to points of interest with small chances of overlap with other commentators. This commentary's contents correspond to its title components.

Something old (and new). Rindermann mentions criticism regarding Buj's (1981) IQ study ('. . . 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 . . . '). This is somewhat reminiscent to posthumous accusations made against Sir Cyril Burt, concerning allegedly non-existent (i.e. fictitious) co-authors, which accusations were wrongful (Fletcher, 1991, pp. 266–276; Hearnshaw, 1979, pp. 242–243; Jensen, 1995). Vinko D. Buj (Croatian, born 1938) can be traced (and asked about study details). He appears in several recent (2006) online papers of Croatian newspapers (e.g. media coverage of Lynn &

DOI: 10.1002/per

Vanhanen, 2006, http://www.slobodnadalmacija.hr/20060721/zadnjastrana01.asp; http://www.slobodnadalmacija.hr/20060701/novosti03.asp; http://vijesti.hrt.hr/arhiv/99/04/07/KUL.html), took his Ph.D. (psychology) from the University of Hamburg (Buj, 1977), with doctoral supervisor Manfred Amelang (Professor Emeritus, University of Heidelberg), a widely known diffential psychologist. Although Buj (1981) is his single entry in the PsychINFO and ISI Web of Science databases and he never again published a similarly large-scaled study, further publications are found in the databases PubMed (Buj, 1983) and Psyndex (Buj, 1990; Buj, Specht, & Zuschlag, 1981). Buj (1981) made clear that his collection of IQ data in 21 European cities ($N = 10.737$) took over 5 years, achieved through a network of native-born collaborators in the respective countries. It is entirely plausible that one nonfamous researcher, apparently a traveller, with good personal contacts abroad, could have carried out this during the mid-1970s. For comparison, the early David Buss amassed data from 37 study sites around the world ($N = 10.047$; Buss, 1989) over a comparable period, starting in 1982, when literally nobody outside the U.S. had e-mail access (Buss, 2003, p. 4; 2004, p. xxi). Richard Lynn (personal communication; February 19, 2007) used Buj's (1981) data (Lynn & Vanhanen, 2002, 2006) because of the totality-of-evidence principle. However, since there is a plethora of other studies, inclusion or exclusion of Buj's data does not impact national IQ estimates. On the whole, the criticism regarding Buj (1981) appears unjustified.

Something new (and old). In the target paper's final section, the isolation of different research paradigms, rooted in different disciplines and fields, is discussed as one of the general problems of cognitive-ability research. To Rindermanns' examples of neglected areas and traditions I add a further one here. There has been a recent revival of research interest into the relations of cognitive ability and suicide risk. Notions about and evidence for an increased suicide risk at the higher intelligence levels can be traced back to 19th-century suicidology (Durkheim, 1897/1997; Masaryk, 1881/1970; Morselli, 1881; see also Goldney & Schioldann, 2002). Positive aggregate-level correlations of intelligence and suicide prevalence have been found in all cross-national studies on this topic (Lester, 2003; Templer, Connelly, Lester, Arikawa, & Mancuso, 2007; Voracek, 2004, 2005b, 2006a), all of them using Lynn and Vanhanen's (2002) national IQ estimates. Within-nation studies have brought mixed evidence, i.e., positive correlations (Kerkhof & Kunst, 1994; Lester, 1995; Voracek, 2005c, 2006d, 2006f, 2006g), negative correlations (Abel & Kruger, 2005) and inconclusive findings (Lester, 1993; Voracek, 2006e). Reasons contributing to the continued interest in geographical studies in this area include: (a) aggregate-level findings such as these normally build up on real individual-level effects (Agerbo, Sterne, & Gunnell, 2007); (b) they are consistent with numerous incidentally ascertained-related findings from suicide research (for detailed reviews, see Voracek, 2004, 2005b); (c) they appear consistent with evolutionary and population genetical theorising about suicide (de Catanzaro, 1981; Voracek, 2006c) and further, similar to Rindermann's key findings, also fit into the sociobiology of life-history traits (i.e. differential *K* theory; Rushton, 2000, 2004). I fully agree with Rindermann that cognitive-ability research would benefit from a broad integration of different research paradigms and lines.

Some things borrowed. Rindermann notes that, for estimating national IQs, applying constant correction factors (2 or 3 IQ points per decade; Lynn & Vanhanen, 2002, 2006) to account for the secular rise in IQ (i.e. the Lynn–Flynn effect), is probably not always correct. This is very likely true. The international pattern of the Lynn–Flynn effect is erratic and poorly understood. Surprising and unsettled cross-national differences in the overall gain rates as well as in the gain ratios for fluid versus crystallised intelligence

DOI: 10.1002/per

measures are evident from the literature (for references, see Voracek, 2006b). However, these and other shortcomings in estimating national IQs actually work against research hypotheses such as those tested by Rindermann (measurement error inevitably attenuates effects). Further, the correlations between national IQ and scores on student assessment studies are quite dramatic (in the .80s or .90s). This is due to the massive data aggregation. Group-level (aggregate) analyses yield larger effects than individual-level analyses (Lubinski & Humphreys, 1996). Accordingly, individual-level correlations of intelligence test results with student assessments are much lower, although still substantial: Rindermann cites two such examples ($r = .47$ and .61). More data on this issue are clearly needed.

And something blue(-penciled). Discussing limitations of the worldwide collection of national IQ (Lynn & Vanhanen, 2006), Rindermann cites two book reviews (Loehlin, 2007; Mackintosh, 2007), noting that some errors in these data have been observed. However, these book reviews relate to another book (Lynn, 2006), not to the source of national IQ estimates.

Methodological Aspects Concerning Rindermann's *g*-factor of International Cognitive-Ability Comparisons

OLIVER WALTER

Leibniz Institute for Science Education, University of Kiel, Germany
walter@ipn.uni-kiel.de

Abstract

It is argued that several methodological aspects concerning the broad definition of literacy and intelligence, the heterogeneous samples, the scaling methodology of international student assessments, the highly aggregated data and the requirements of higher-order factor analysis provide sound alternative explanations to Heiner Rindermann's hypothesis that literacy could be subsumed under the intelligence construct. Copyright © 2007 John Wiley & Sons, Ltd.

Rindermann (this issue) identifies a 'strong *g*-factor of differences between nations' on the basis of correlations between scales of several international student assessments (e.g. PISA) and intelligence tests. He argues that this *g*-factor indicates a general cognitive ability and hypothesises that literacy could be subsumed under the intelligence construct. In my opinion, Rindermann's analyses and interpretations suffer from several theoretical and methodological misconceptions some of which have been discussed elsewhere (see Baumert et al., 2007; Prenzel et al., 2007). Here I will focus on Rindermann's construction and interpretation of *g* and will argue that his findings can alternatively be explained by several methodological aspects.

First, in international student assessments, mathematical, reading and scientific literacy are broadly defined constructs which include many more or less different abilities and

DOI: 10.1002/per

skills. The same is true for intelligence because it also covers a wide range of different abilities. Therefore, it is obvious that some cognitive processes which are essential for solving items in student assessment tests will also be needed for solving items in intelligence tests. The relationship between student achievement and intelligence cannot be doubted since Binet constructed his intelligence tests to identify children who need special education, and since school achievement is used for the validation of intelligence tests. Hence, positive correlations between data from intelligence tests and data from tests measuring student achievement can be expected and are generally found. However, positive correlations are not sufficient to conclude that intelligence tasks and items in international student assessments are indistinguishable, that cognitive demands across different scales are homogeneous or that 'the ''literacy'' concept could be included in the historically older intelligence concept'. For example, body size and body weight are positively correlated, but neither do we think of them as the same nor do we suggest that one could be subsumed under the other. Furthermore, goodness of fit tests have shown that intelligence measured by a subtest of a German intelligence test (Heller & Perleth, 2000), which is regularly administered as a national option in PISA, is distinguishable from the four literacy scales in PISA 2003 (see Prenzel et al., 2007). In addition, theoretically important differences have been found between groups of students, e.g. boys and girls, on some of these scales, but not on others. Both findings indicate that they measure different constructs in fact, although these scales are highly correlated.

Second, Rindermann's question as to why these scales are so highly correlated can be answered by looking closely at several methodological aspects of international student assessments. The large samples used in such studies cover a wide range of cognitive abilities between students and countries all over the world. Although heterogeneity does not necessarily lead to high correlations, it very likely promotes them if there is a linear relationship between variables. Another aspect that promotes high correlations is the *direct* estimation of latent correlations on the group level in TIMSS and PISA (see Mislevy, Beaton, Kaplan, & Sheehan, 1992; Walter, 2005, for a description of this methodology). Because such correlations are not attenuated by individual measurement errors, they are usually higher than correlations between traditional scale values. Finally, Rindermann's aggregation of data on the country level eliminates theoretically important differences within countries and raises the correlations even further. Therefore, his findings of high positive correlations between literacy scales and intelligence tests are not surprising.

Third, Rindermann's subsequent factoring of these high correlations lead to the extraction of only one factor, i.e. *g*. Because these correlations are correlations between scales and not between raw data, he has actually conducted higher-order factor analyses. It is well-known from factor-analytic research that the number of higher-order factors is strongly dependent on the design of the studies analysed. For extracting more than one higher-order factor it is necessary to include a considerable number of scales which measure a great variety of constructs. Since these methodological requirements are not met in the majority of studies, higher-order factor analyses very often lead to the extraction of only one factor (see Carroll, 1993, for a more detailed discussion and empirical findings). In the target paper, the methodological preconditions for extracting more than one higher-order factor are not met, either, because Rindermann factored correlations between scales measuring only four or five global constructs which are positively correlated, measured in very heterogeneous samples and aggregated on the country level. Hence, the extraction of *g* just seems to be a consequence of these methodological preconditions, but not a theoretically important result.

DOI: 10.1002/per

For the same reason, it is very doubtful that Rindermann's analysis is a powerful test of his hypothesis that 'cross-cultural correlations between different scales of student cognitive ability [are] high enough to justify the assumption of a strong *g*-factor'. Moreover, the methodological aspects that lead to the extraction of only one factor let me further doubt whether this hypothesis could ever be rejected with the methods use. Although student achievement and intelligence do not show relationships *only* for methodological reasons, the multitude of methodological aspects, which contribute to high correlations and the extraction of only one factor, has to be regarded as a very likely alternative explanation to Rindermann's theoretical hypotheses. Finally, the extraction of a general factor does not support the hypothesis of homogeneous cognitive demands. Although latent variables like *g* can explain differences *between* persons, they do not indicate what is going on *within* persons and their 'minds' (see Borsboom & Dolan, 2006).

In conclusion it is my opinion that Rindermann's extraction of *g* is not a theoretically significant result, but rather a consequence of the methodological aspects I have described. Therefore, it does not support his hypothesis that literacy could be subsumed under the intelligence construct or that cognitive demands are homogeneous in international student assessments and intelligence tests. Future research should look at these methodological issues more closely before drawing such strong theoretical conclusions.

Percentages of Children Living in Poverty Determine IQ Averages of Nations

VOLKMAR WEISS

Leipzig, Germany
volkmar-weiss@t-online.de

Abstract

By comparing three bodies of independently collected data sets, Lynn–Vanhanen-IQ, PISA-IQ and children poverty percentages, we have evidence of a downward and hence dysgenic trend in a number of nations, reaching up to 6 points within one generation and even higher losses in Latin American countries. Copyright © 2007 John Wiley & Sons, Ltd.

Already in the last quarter of the 19th century the decrease of birth rates in the upper stratum led to the assumption of a threat of an accompanying decrease of the average giftedness of a nation (Blacker, 1952). But contrary to all such expectations cognitive test scores were rising over many decades. For a geneticist (Weiss, 1992) it seems clear that—in analogy to the acceleration of body height—such a rise could only be a rise in phenotypic values and not in genotypic ones. But in view of the Flynn effect the argument that a dysgenic development was imminent seemed to be ridiculous (Neisser, 1998). Unifying the results of international educational research and differential psychology into a most plausible estimate of respective national IQ, this is the merit of Rindermann's target paper.

DOI: 10.1002/per

To measure and hence prove a dysgenic trend of IQ is extremely difficult. If the sea level is rising or falling the shore line of an island remains fixed. The IQ, however, is a relative measure always gauged to the median of a reference population which is only in an idealised population and distribution of test scores identical with the arithmetic mean of a bell curve. Lynn and Vanhanen (we refer here only to 2004) chose the median of the United Kingdom of 1979 as the score 100 of their 'Greenwich IQ'. But no country can claim to be resistant to any change. In the 2000 and 2003 PISA studies, 500 and 100 are chosen as mean and standard deviation, with the effect that in 2003 by the first time inclusion of Turkey into the sample of reference the average 'PISA-IQ' of Germany and others rose without their contributing anything to such an effect. Educational politicians in industrialised countries could even be more proud of their nations rising IQ if Brazil and other third-world nations would be included into the sample on which basis the general mean is calculated, and even prouder if the nations would be weighted by the total numbers of school children in the respective countries (Weiss, 2006).

In order to minimise such methodological pitfalls I set in the following the (by Rindermann) corrected arithmetic mean of the PISA scores of Netherland (528), New Zealand (525) and the United Kingdom (528) as IQ 100; 15, corresponding to 527; 100. In 2000 and 2003, PISA subjects were of an age of 15 years. It is a pity that Lynn and Vanhanen do not give the average birth years of their data sets but it can be assumed that the subjects in their collection came from the parental generation of the PISA subjects.

In the study 'Child poverty in rich countries 2005' the poverty threshold is defined as the percentage of children living in households with incomes below 50% of the national median income. The percentage of children living in poverty could be high, because many children are born to the poor ore because the well-to-do have relatively less children. In (former West) Germany, for example, more than 40% of women with an academic degree remain childless (Weiss, 2002). 'The Report series has regularly shown, there is a close correlation between growing up in poverty and the likelihood of educational under-achievement, poor health, teenage pregnancy, substance abuse, criminal and antisocial behaviour, low pay, unemployment and long-term welfare dependence. . . . Such problems are associated with, but not necessarily caused by, low income (for example, low levels of parental education or parental skills)' (UNICEF, 2005, p. 6).

A high percentage of children in poverty could be a strong hint to a dysgenic trend in the respective country, a small percentage a hint to an eugenic trend. Indeed, the eight richest countries (Denmark, Finland, Norway, Sweden, Switzerland, Czech Republic, France) in the 'Child Poverty League' (UNICEF, 2005, p. 4)—with no more than 7.5% of their children living in poverty—have on average no IQ decrease; Finland, where only 2.8% children live in poverty, even an IQ-increase of 6 points. But Germany (loss of 6 points down to a mean IQ of 96), Italy (children poverty 16.6%; IQ loss of 7 points) and Mexico (children poverty 27.7%; IQ loss of 17 points) exhibit a clearly dysgenic trend.

Now we bring these poverty percentages in context with the percentage of children who got a PISA-IQ of 88 and lower: The 15 countries that have a percentage of low-IQ children below average (Denmark, Finland, Norway, Sweden, Switzerland, Czech Republic, France, Belgium, Netherlands, Austria, Japan, Australia, Canada, Ireland, New Zealand) have on average the same mean IQ as given by Lynn and Vanhanen. However, the eight countries with an above average percentage of dull children (Hungary, Germany, Greece, Poland, Spain, Italy, USA, Mexico; Luxemburg excluded because of its smallness) have a mean IQ loss of 6 points. In this way, comparing three bodies of independently collected data, Lynn–Vanhanen-IQ, PISA-IQ and children poverty percentages, we have evidence

DOI: 10.1002/per

for eugenic and dysgenic trends on a national scale, reaching up to 6 points within one generation. Maybe the trend would be even more visible if the Lynn–Vanhanen-IQ would be uniformly scaled to the year of birth 1960.

Australia and Canada show an IQ-increase despite high percentages of children poverty. Maybe a dysgenic trend is checked by selective above average IQ immigration.

Very disquieting are IQ changes in Latin American and third world countries: Argentina Lynn and Vanhanen IQ 96 (PISA-IQ 77); Brazil 87 (68); Chile 93 (81); Indonesia 89 (72); Mexico 87 (70); Peru 90 (63); Uruguay 96 (79).

The downward trend has an easy explanation: In Brazil, for example, the 2.5% of women living in the top income group had less than 2.0 children already in 1970 (Wood and de Carvalho, 1988, p. 191). However, in the four poorest strata comprimising 48.5% of the population, women had 7.4 children. Consequently, their share of the population grew to 58% in 2000, the share of the top income group dwindled to 1.4%.

What is the National *g*-Factor?

JELTE M. WICHERTS[1] and OLIVER WILHELM[2]

[1]*Department of Psychology, University of Amsterdam, The Netherlands*
[2]*Institute for Educational Progress, Humboldt University Berlin, Germany*
J.M.Wicherts@uva.nl, oliver.wilhelm@rz.hu-berlin.de

Abstract

Rindermann correlated the national averages of several student assessment studies and 'national IQ' estimates and proposes that these variables are all indicators of a common cognitive ability at the macro-social level, which he denotes the national g-factor. We argue that Rindermann oversimplifies issues of individual differences and applies inappropriate statistical analyses. Therefore, we refute his conclusions. Copyright © 2007 John Wiley & Sons, Ltd.

Rindermann (this issue) implicitly assumes that his national *g*-factor is homogenous with the *g*-factor at the individual level. We argue that Rindermann has failed to establish that the nature and causes of this national *g*-factor are similar (or identical) to the nature and causes of the individual *g-factor* (as for example in Carroll's, 1993, widely accepted model).

Nature of the national g-factor. Probably only a few educational or intelligence researchers would have issues with the idea that tasks from large-scale educational studies share critical attributes with intelligence tasks. Nonetheless, most ability researchers will likely dismiss the notion that two different ability tasks such as Raven's Standard Progressive Matrices and the Reading Literacy test from PISA measure the same underlying ability. There is no serious doubt about the necessity of more specific lower order or nested factors to adequately account for covariances between ability tests (Carroll, 1993). If, for pragmatic reasons, only single indicators rather than a comprehensive intelligence test battery can be used, such measures should be decontextualised reasoning measures like the Raven's tests because there is little to no contribution to performance

DOI: 10.1002/per

from intelligence factors other than *g* (at least in the West) (Gustafsson, 1984; Wilhelm, 2005). The literature on the factor-analytical structure of inter-individual differences in cognitive ability is neglected when Rindermann factor analyses the average scores of distinct achievement indicators across countries as indicators of a single underlying ability.

Rindermann (this issue) discusses average scores on the level of nations in terms of notions drawn from studies at the level of individuals. However, he fails to establish a bridge between the individual and the national level. Multi-level data require the use of multi-level analyses (Muthén, 1991). Unfortunately, Rindermann uses nation's averages as his input and fits a single factor model at the level of nations' averages, without verifying that test scores show comparable within-country factor structures across nations. The ecological correlations are indeed high, but it has long been known that ecological correlations may lead to results that are inconsistent with inter-individual correlations (Robinson, 1950).

For a comparison of test scores across nations to be meaningful, we have to establish that the tests measure the same underlying construct across nations and that the tests are not biased with respect to nations. Moreover, we have to establish that the underlying construct we are looking at across nations is indeed *g* and not another of the lower-order factors that are known to play a role in cognitive test performance (Carroll, 1993). The starting point would be to study measurement invariance (Mellenbergh, 1989) across nations. Unfortunately, Rindermann does not discuss the issue of measurement invariance, he does not fit the relevant multi-level factor models (see Muthén, 1991), he does not consider model fit, nor does he study alternative factor models to the structure of aggregated data. Thus, Rindermann's analyses failed to show that national differences in average achievement and IQ test performance are due to national differences in average inter-individual *g*.

Causes of the national g-factor. Even if the factor-analytic results were trustworthy—and we must stress that they are not—we still need to consider what explains differences in national *g*. We do not know whether national differences in ability test performance have a genetic basis. This would require the finding of genes that account for intelligence and that differ in distribution across the globe (because cross-national twins do not exist). Lynn (2006) and Rushton (2000) pose evolutionary theories related to race, but archaeological findings are strongly at odds with these theories (MacEachern, 2006).

More importantly, a comparison of nations across the globe is fraught with many highly relevant confounds at the macro-social level. Many of these confounds are known or at least suspected to have strong effects on average cognitive-ability test scores, which have risen substantially in developed countries over the course of the 20th century (i.e. the Flynn Effect; Neisser, 1998). Health, nutrition, education, urbanisation, trends towards smaller families, and the introduction of computers have all been proposed as causes for the Flynn Effect. However, developing countries like those in sub-Saharan Africa have not seen such developments. Wicherts (2007) has shown that estimates of national IQ (from Lynn & Vanhanen, 2002) correlate highly with basically all variables that have been proposed to have caused the Flynn Effect, such as secondary enrollment ratio (.78), pupil-to-teacher ratio (−.72), the number of PCs per 1000 persons (.66), fertility rate (−.86), urbanisation (.67), general health as expressed in the child mortality rate (−.81), and nutrition as expressed in the amount of proteins in *g* per day *per capita* (.76). This begs the question whether the national *g*-factor is indeed something that looks like *g* at the individual level, because on the basis of such strong correlations, national *g* looks suspiciously similar to the developmental status of countries.

DOI: 10.1002/per

We must add that the results by Rindermann are further compromised by many systematic flaws in Lynn and Vanhanen's (2002, 2006) IQ data. For example, the IQ values for African countries are consistently too low (Wicherts, 2007), which artificially inflate the correlations reported by Rindermann.

Conclusion. We think that Rindermanns' paper nicely illustrates three fundamental issues in the study of group differences. First, in order to understand group differences in some construct it is necessary to understand the nature of individual differences in this construct and we doubt Rindermann fully understood individual differences in the achievement data he reanalysed. Second, when investigating group differences the use of adequate statistical tools (Mellenbergh, 1989; Muthén, 1991) is crucial, and unfortunately Rindermann fell short of using this set of appropriate tools. Third, when making controversial statements about the differences in national *g* through reanalysis of available data, these data must fit the problem and must be of unequivocal quality. The PISA, TIMSS and PIRLS data used by Rindermann are of very good quality but don't fit to the problem. The IQ data used by Rindermann might fit to the problem, but are of poor quality.

DOI: 10.1002/per