MANKIND QUARTERLY
Mankind Quarterly, 66(2), 226-242 2025 Winter Edition
ISSN: 0025-2344

International Student Assessments Measure Cognitive Ability: A Response to Rutkowski et al. (2024)

Heiner Rindermann*

Abstract

Rutkowski et al. (2024) questioned whether international comparative student assessments (e.g., PISA, TIMSS, PIRLS) also measure intelligence. In addition, they framed the research of other scholars as being on the side of political evil (e.g., the "holocaust") and accused them of having sinister intentions (e.g., "eugenics"). To determine what a diagnostic instrument measures, one can choose a content-cognitive-psychological or an empirical-correlative approach. To begin with, however, it is necessary to define terms and constructs. We define intelligence as the ability to think and the broader cognitive ability (or cognitive competence) as the ability to think, the disposal of knowledge and its use in an understanding way. Analyses show that the cognitive demands of the tasks in intelligence tests and student assessment tests are similar and can be solved using similar cognitive processes and strategies (e.g., finding information and reasoning). Correlative and factor analytic analyses at different data levels show a high empirical similarity and a strong common factor. The causal factors for individual development and for individual, national and historical differences in IQs and ILSA (international large-scale assessment) are also similar. Rather than distinguishing between intelligence tests and student assessments, it is recommended to distinguish between thinking and knowledge. Finally, the introduction of political ideology is not helpful; science is damaged when hatred and agitation are spread. The Nazis, for example, also denigrated intelligence research, which from their perspective was part of a Jewish-modernist ideology.

Keywords: Student achievement; Intelligence; Cognitive ability; International large-scale assessment (ILSA)

1 Terms and concepts: Intelligence as the ability to think vs. knowledge

Rutkowski et al. (2024) questioned whether international comparative student assessment studies (PISA, TIMSS, PIRLS and regional studies) also measure intelligence.[1] When considering what a test does or does not measure, one should first establish a definition of the psychological constructs in question, then analyze the tests and compare the results with those constructs (content-cognitive-psychological analysis). By failing to provide definitions of its central terms and constructs, the paper by Rutkowski et al. (2024) leaves the meaning of "intelligence", "learning" and "student assessment" indeterminate, thereby undermining the significance of its purported contribution. One cannot claim that a test fails to measure something ("A") but measures something else ("B") if neither "A" nor "B" has been defined. In this paper, an attempt is made to rationally reconstruct the approach of Rutkowski et al., beginning with proposed definitions and then testing their claims.

Intelligence is understood here as the ability to think, a rather knowledge-reduced mental capacity that is ideally free of specific knowledge (Rindermann, 2018, p. 43). Intelligence comprises:

*Department of Psychology, Chemnitz University of Technology, Germany

[1] PISA: Programme for International Student Assessment; TIMSS: Trends in International Mathematics and Science Study; PIRLS: Progress in International Reading Literacy Study.

- *Problem solving*: to solve new problems by thinking rather than simple knowledge recall,
- *Reasoning*: to infer (to conclude and reason, to draw inductive and deductive-logical conclusions including finding patterns in information, to correctly generalize, to apply rules for new examples and to solve syllogisms),
- *Abstract thinking*: to categorize, to form concepts, to process abstract information in the form of verbal and numerical symbols, in the form of abstract figures and in the form of general rules,
- *Understanding*: to recognize and construct relationships, structures, contexts and meaning, to have insight.[2]

The complementary term to intelligence (i.e., the concept with which it is most appropriately contrasted) is not student achievement or assessment — which is undefined in terms of content and could include areas such as sports — but rather knowledge. *Knowledge* is the possession of true and relevant information: statements that accurately describe reality (e.g., the Earth is round, the Romans lived around 2,000 years ago, tigers and whales are mammals, etc.).[3] *Cognitive ability* comprises the ability to think (intelligence), knowledge, and the intelligent use of this knowledge. "Cognitive competence" can be used interchangeably.[4] Knowledge is not independent of intelligence because thinking abilities are essential for acquiring and using knowledge: New information must be detected, connected to existing knowledge in long-term memory, integrated into the existing network, and linked appropriately. This process also requires the formation of hierarchical categories and abstraction. In the process of thinking, knowledge must be purposefully retrieved and often newly combined.

Student achievement (or educational achievement) is primarily the recognizable performance of students in school examinations that are graded by teachers (Rindermann, 2018, p. 51). Student achievement can be measured more objectively via student achievement or assessment tests. Compared to intelligence tests, they should rather measure the knowledge acquired at school. Rutkowski et al. suggest that student assessment tests measure "learning". This is not correct, because "learning" primarily refers to a process: the action or the process of acquiring knowledge. It would be better to say that they measure the *result* of learning.

So, what psychological traits do the given student assessment tests measure? To answer this, I will briefly look at the tasks, cognitive demands and processes as well as empirical correlations.[5]

2 Content validity: Analyses of tasks and cognitive processes

In a systematic rating study, eight PISA tasks and four TIMSS tasks were assessed by 68 persons (teachers and psychology students; Rindermann & Baumeister, 2015). They rated the content and the cognitive competences and processes required for solving the tasks. Intelligence was considered more important than

[2] This is similar to Gottfredson's (1997, p. 13) definition: "Intelligence is a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience."

[3] A colleague pointed out that "knowledge" may not be true and that the truth value of knowledge may change over time. This is an important objection to help clarify the term. An older philosophical-normative and a younger psychological-descriptive concept can be distinguished: In the first concept, knowledge stands for intellectual contents that are subjectively and objectively certain, in contrast to mere opinion and belief. In the second concept, knowledge stands for any mental content, everything that is stored in memory. The first also fits in with the concept of knowledge that is transmitted at school and that is to be assessed through classic student achievement tests. The second fits better with a cognitive psychology perspective. However, both concepts can be understood as the opposite of intelligence as the ability to think. In the context of student achievement studies, the first concept makes more sense. Knowledge, in the sense of this second concept, is best understood as the storage of any information.

[4] There are further cognitive constructs such as mental speed, memory, problem solving, creativity etc., which are for the purpose of this paper less relevant.

[5] One reviewer mentioned that correlations are always empirical, which, of course, is correct. Nevertheless, the term "empirical" is used here to emphasize the complementarity of empirical-statistical and qualitative analyses.

knowledge for solving nine of twelve tasks (or alternatively measured, six out of eight tasks). If a task is examined in more detail, e.g., “Lake Chad” from PISA 2000[6], it becomes obvious that it is a general cognitive ability rather than knowledge that is being measured (Rindermann, 2018, p. 51ff): It is necessary to read the given text, figures and numbers, retrieve and understand information, conclude/infer/reason and take the perspective of others. The retrieval of specific knowledge acquired in class rarely plays a role (Rindermann & Baumeister, 2015).

In any case, it is clear that student achievement tests are not school knowledge tests, nor were they ever intended to be. PISA and PIRLS, in particular, were designed to measure “literacy”, conceptualized as the ability to solve cognitive tasks encountered in school, at work, and in everyday life in modern societies.[7] This is *not* a bad idea, and it is not criticized here. Rather, it makes both the construct and the tests relevant to the cognitive challenges of modernity, while at the same time positioning them further from school knowledge and closer to intelligence. The intention behind international school assessments should also be mentioned. As the PISA organizers state:

> *“PISA goes beyond assessing whether students can reproduce what they have learned in school. To do well in PISA, students have to be able to extrapolate from what they know, think across the boundaries of subject-matter disciplines, apply their knowledge creatively in novel situations and demonstrate effective learning strategies.” (OECD, 2019, p. 5)*

Interestingly, this is also a useful definition of an important aspect of intelligence. The PISA researchers seem to assume that one of the primary purposes of schooling is to make children better thinkers. This assumption is supported by empirical evidence suggesting that IQ rises as a result of school education (Ritchie & Tucker-Drob, 2018).

3 Empirical correlations

Rutkowski et al. (2024, p. 3) wrote: “ILSAs have never been validated as measures of intelligence at the individual or population level.” Additionally, they wrote (p. 5): “Each dimension serves as an aggregated measure of learning in the respective content domain.” Neither statement is consistent with the empirical literature (“validated” is understood as “investigated and confirmed”). Given the extensive research, it seems odd to make such a claim (“never been”): There are many studies at different levels of data that Rutkowski et al. failed to take into account.

(1) Correlational and factor analytical studies at the *individual* level including or excluding intelligence tests.

(2) Correlational and factor analytical studies at the *population* level including or excluding intelligence tests.

(1) A correlation study by Brunner (2008), published in the same journal *Learning and Individual Differences* as Rutkowski et al. (2024), provided initial evidence: The average manifest and latent correlations between PISA-Reading and PISA-Math were smaller than their average correlations with psychometric CogAT-scales (Cognitive Abilities Test; averaged manifest $r_{\text{PR-M}} = .45$ vs. $r_{\text{P-C}} = .47$, latent $r_{\text{PR-M}} = .80$ vs. $r_{\text{P-C}} = .86$).

In another study by Pokropek et al. (2022), the correlations among the PISA dimensions of reading, mathematics, and science at the latent level averaged $r = .86$. The correlation between these three PISA dimensions and Raven’s figural intelligence test was $r = .73$ (also latent). Here was a specific PISA factor, which perhaps does not represent student achievement, but rather a PISA method factor. In any case, there were no content-specific PISA factors:

[6] Sample collection of tasks: https://nces.ed.gov/surveys/pisa/pdf/items2_reading.pdf

[7] “PISA measures 15-year-olds’ ability to use their reading, mathematics and science knowledge and skills to meet real-life challenges.” (OECD, 2024)

> *"The domain-specific factors are not reliable enough to be interpreted meaningfully. They lie somewhere between unreliable measures of domain-specific abilities and nuisance factors reflecting measurement error." (Pokropek et al., 2022, p. 1121)*

This means that empirically, the PISA scales do not clearly measure reading, mathematics or science, but rather a general ability that sometimes even correlates more strongly with figural intelligence than with any other PISA scale. As a consequence, discriminant validity is not given. Of course, it is then possible to calculate a total PISA value, as Altinok and Diebolt (2024) have done, for example. This also means that it makes sense to use a summary value for cognitive human capital in economic, political, and sociological analyses, because the specific effects are rather small — although they do exist for some theoretically justified questions, e.g., mathematics vs. language tilt (Becker et al., 2022).

Independent of ILSAs (international large-scale assessments), Deary et al. (2007) reported a correlation between the central compulsory English school exam GCSE (General Certificate of Secondary Education for 15- to 16-year-old students) and the CAT (or CogAT, Cognitive Abilities Test) of $r = .69$ (latent $r_l = .81$), with the CAT figural scale alone $r = .66$. The correlations are probably underestimated as there was a four to five year gap between CAT and GCSE (11 year olds and 15 to 16 year olds). This high correlation means that, regardless of the test construction specifics of the ILSAs, student achievement itself is highly correlated with intelligence. Many other studies using conventional student achievement and intelligence tests came to similar results (e.g., Kaufman et al., 2012; Zaboski et al., 2018). As early as 1954, Coleman and Cureton spoke of a "jangle fallacy" if intelligence and student achievement tests were given different names. They replicated with a correlation of $r = .83$ to $.84$ (between the Otis Quick-Scoring Test Beta and the Stanford Achievement Test) and with their term "jangle fallacy", what Kelley (1927) had already found and formulated in the 1920s with other samples (in his case between intelligence and student achievement tests corrected for measurement error $r_l = .90$; p. 196). Kelley put it aptly (1927, p. 64):

> *"Contaminating to clear thinking is the use of two separate words or expressions covering in fact the same basic situation, but sounding different, as though they were in truth different. The doing of this latter the writer will call the 'jangle' fallacy. 'Achievement' and 'intelligence' sound as though they were different; they have different 'jangles', and thus we treat them as though they were different in truth."*

Jussim (2020) calls it briefly and succinctly in his Orwelexicon, picking up on the current political situation in academia and media: "IQaphobia: Fear of measuring intelligence because one believes that only Nazis and Eugenicists do that."

(2) There are numerous studies that show a high correlation between student assessments and intelligence tests at the population level (e.g., Boman, 2023, $r = .766$ between PISA-2018 and national IQs of Lynn & Becker, 2019, in 77 countries). This leads to a very strong G factor (see Figure 1).

This correlation is also independent of which student achievement and intelligence test collections are used. For example, if we take the latest World Bank ILSA collection (Altinok & Diebolt, 2024) and the intelligence test collection by Lynn and Becker (2019; only measured data), the result is a correlation of $r = .77$ ($N = 108$ country means). Warne (2023, his Table 1) reported correlations with similar but somewhat older student achievement data sets in the range of $r = .79$ to $.83$. Within Germany, at the level of 16 states, the correlation between PISA and intelligence is $r = .71$ (Rindermann, 2024).

4 Summary on what international student assessments measure

Applying a narrow concept of intelligence (as thinking) and a more operational concept of student achievement (relatively knowledge-intensive performance of students in school exams) and based on analyses of content,

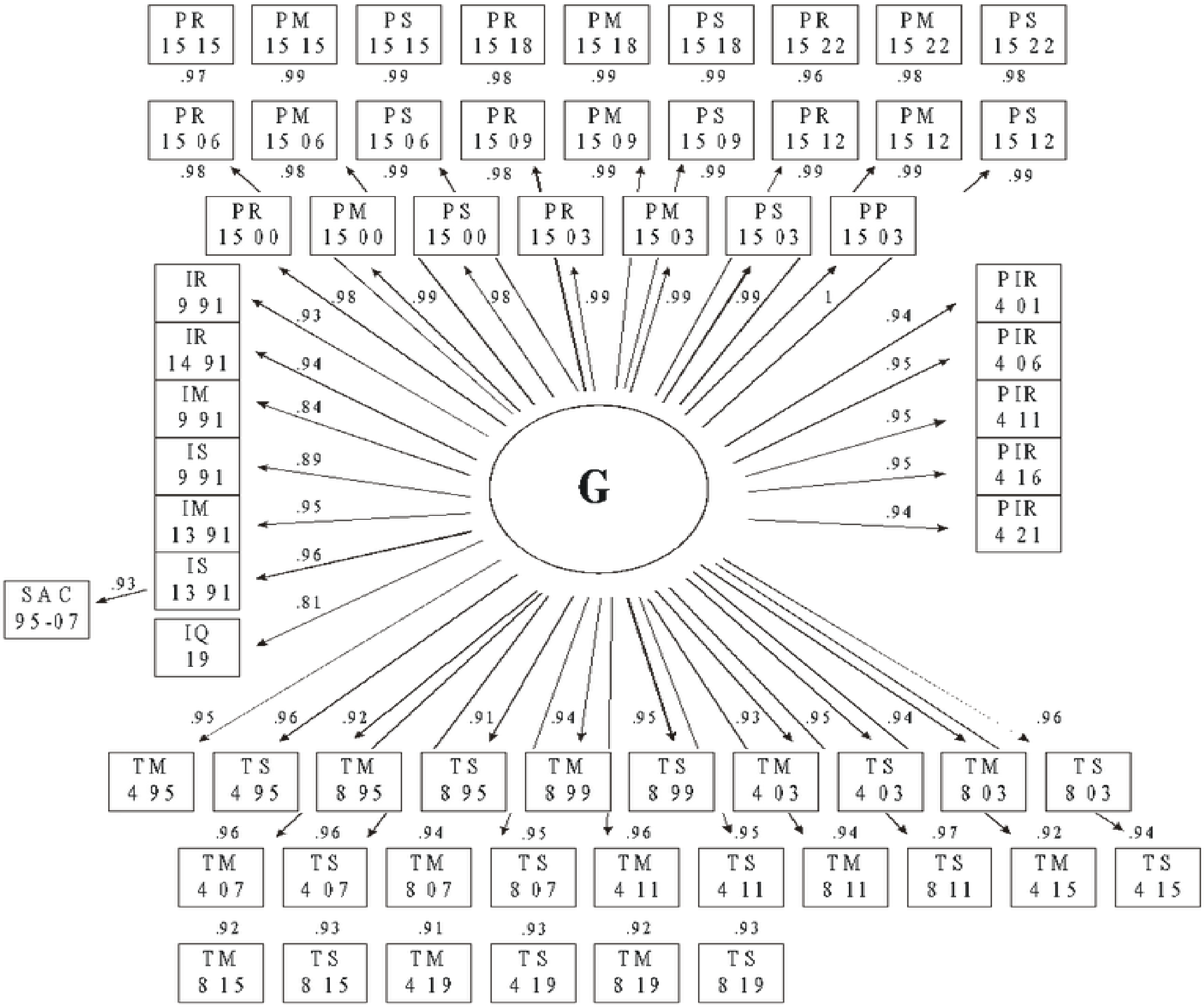

Figure 1: *G* factor of international differences in 64 student achievement and intelligence test scales (PR: PISA Reading, PM: Math, PS: Science, PP: Problem solving, 15 year old students and survey year; PIR: PIRLS, reading in 4th grade and survey year; TM: TIMSS Math, TS: TIMSS Science, 4th or 8th grade and survey year; IR: IEA Reading, IM: IAEP Math, IS: IAEP Science, 9 or 13 year old students and survey year; SAC: SACMEQ, waves I, II and III from 1995 to 2007; IQ: Lynn & Becker measured data published 2019; on arrows loadings max. 1, average loading is $\lambda = .95$, based on uncorrected national means; from Rindermann, 2018, p. 95, updated)

of cognitive processes and of empirical relations, student assessment tests also measure intelligence. While intelligence can be clearly defined, student achievement is a psychologically less clear-cut construct. For theoretical reasons, it would be better to distinguish between thinking and knowledge than between intelligence and student achievement.

However, it should be mentioned that in addition to this rather narrow concept of intelligence proposed here, there are others that (I believe) are even more widespread in research. If, for example, we take Cattell's model (1987/1971), which distinguishes between fluid intelligence (pure reasoning) and crystallized intelligence (application of knowledge, explicitly including knowledge), then student assessment tests and what they measure would certainly be part of the intelligence construct. Many widely used intelligence tests include knowledge-based questions or tasks that rely on content typically learned in school, such as fractions or geometry, as seen, for example, in the CogAT (Lohman & Hagen, 2002). These tasks cannot be solved without prior knowledge acquired at school. Another example are tests from the Wechsler tradition (WISC and WAIS).[8]

[8] WISC: Wechsler Intelligence Scale for Children; WAIS: Wechsler Adult Intelligence Scale.

Although the aspect of content validity is in the foreground and is the decisive one, there are also other approaches. In another view of "what a variable measures", a test or variable measures everything with which it correlates, even if it has a different content. For example, it is possible to "measure" gender with hair length or weight with height. Similarly, to "measure" ability levels in math, variables such as reading, gender, parental education, migration background, class and school average can be used. Such correlations are used in international student assessment studies in particular, for example, to extrapolate ability scores in reading and maths for students who only completed science tasks (e.g., NEPS, n.d.).[9]

That is exactly what successful companies do: Everyone who buys dog food is also likely to be interested in collars. That is applied thinking, intelligence in practice. Nevertheless, it would be better, in such cases not to speak about "measurement" but about "estimation". Although they are correlated, we should continue to distinguish thinking from knowledge. The analysis of *content* and *cognitive processes* cannot be replaced by purely correlational approaches (for example, if one were to consider only correlations with any characteristics).

This is *not* to suggest that student achievement tests are inherently bad. Rather, they are not strictly knowledge-based and bear strong similarities to intelligence tests.[10]

5 Common causes for results in IQ tests and ILSAs

Another argument in favour of intelligence tests and student assessments measuring similar or identical constructs is that the results in them depend on similar or identical causes:

(1) *Genetic factors* most probably have a global rather than a specific influence on cognitive ability (e.g., Bartels et al., 2002; Haworth et al., 2008; Shakeshaft et al., 2013).

(2) *Basic cognitive competences* like attention control, mental speed and working memory exert a rather global influence on more complex cognitive competences (e.g., Jensen, 2006; Schneider & Niklas, 2017).

(3) *Environmental factors* such as physical, cultural-familial and school environment show also global influence (Hattie, 2023; Rindermann, 2018; Steinberg, 1996).

(4) During development, there are *positive interactions between several subsystems*, for example, between reasoning and knowledge, interest and competence (Harackiewicz et al., 2008; Rindermann et al., 2010; Vu et al., 2024).

(5) Similar cognitive processes are required for *solving tasks* from different tests and scales, for example, attention control, concept formation, inductive and deductive reasoning, knowledge retrieval and knowledge application (Burgoyne & Engle, 2020; Kovacs & Conway, 2019; see also above Rindermann & Baumeister, 2015).

(6) In a *test situation*, *similar personality traits* are important for being successful in different tasks such as diligence, effort and low test anxiety (e.g., Mammadov, 2022; Stanek & Ones, 2023).

[9] The estimation methods in more detail in quotations, e.g., for PISA: "In practice, plausible values are generated through multiple imputations based upon pupils' answers to the subset of test questions they were randomly assigned and their responses to the background questionnaires." (www.oecd.org/pisa/data/httpoecdorgpisadatabase-instructions.htm). For the German NEPS (National Educational Panel Study), there exists a list of variables for calculating plausible values for reading (Scharl et al., 2020, Table 6, p. 18): "Mathematics competence [s], Procedural metacognition (mathematics) [s], Reading speed [s], Procedural meta-cognition (reading) [s], Reading activity (in hours): off the job [s], Reading activity (in hours): on the job [s], Age at wave 3 [s], Migration background of test target [s], Female [s], ISEI-08 (Socio-economic status) [p], Employment status [p], ISCED-97 (Educational level) [p], Years of education [p]." ([p] and [s] are added by HR) This means that parental characteristics [p], such as SES, and child characteristics [s], such as migration background, which theoretically have nothing to do with an ability (no content validity), are used as variables correlated with abilities to estimate abilities.

[10] We have not addressed here that the claim that student achievement is unrelated to intelligence may also be strategic in nature. A testing organization might make such a claim to avoid politically motivated resistance to testing, a strategy that would be instrumentally rational. There is no indication that Rutkowski et al. are adopting such a strategic position.

All these factors additionally confirm what was found in content-related cognitive analyses: What intelligence and student achievement tests measure is similar. They behave similarly, and a sum value can be used for predictive or causal analyses.

Country data represent averaged individual data. This means that intelligence and student achievement at the country level can also be traced back to similar factors. In addition, there are factors at the country level such as wealth, politics and, above all, culture, which also have a global rather than a specific effect on intelligence and student achievement (e.g., Rindermann, 2025, Table 3).

5.1 Correlation with effort

Rutkowski et al. (2024, p. 5f., with reference to OECD) assume that international differences in ILSAs depend on invested effort:

> *"These findings indicate that reported achievement represents performance on the test but it may not reflect actual proficiency. This point especially undermines the previously reviewed literature, which views ILSA scores as infallible measures of innate intelligence that are hardly or not at all influenced by motivation."*

Apart from the fact that there is no one who takes the position that "ILSA scores [are] infallible measures of innate intelligence"[11], the sentence contains an (indeed plausible and) empirically testable statement: effort increases results in ILSA scores.

At the individual data level, there is only a slight positive effect of effort on cognitive performance ($d = 0.17$; Bates & Gignac, 2022). According to Cohen's (1992) usual interpretation standards, this is below a small effect ($d = 0.20$). And what is the situation at the international level? The data can be taken up and correlated with the PISA results (average at the country level).[12] The PISA 2018 figure provides country averages for three effort variables:

(1) Average effort invested in the PISA test,

(2) Average effort students would have invested in the PISA test if scores on the test were going to be counted in their school marks,

(3) Percentage of students indicating that they invested less effort in the PISA test than if their scores were going to be counted in their school marks.

Variable (1) stands for effort invested, variable (3) for low effort (large numbers mean low effort resp. reduction in effort) and variable (2) is an unusable conjecture. The results are these: (1) More effort means lower PISA 2018 ability results ($r = -.43$, $N = 79$ country means) and (3) reduced effort means higher PISA 2018 results ($r = +.54$, $N = 79$).[13] The more effort students in a country report, the lower are the results in this country. Motivation-related self-assessments are not comparable across countries due to frame of reference problems. As is often said, "A beautiful theory killed by ugly facts." But this is science and that is how it should be done. Now, the relevant thing here is why Rutkowski et al. put forward a theory (effort→ILSA country results), but did not test it. *This is not science*. The result is exactly the opposite of what they assumed.

[11] Intelligence tests measure phenotypic, observable intelligence, not genotypic intelligence. Polygenic scores (as correlated with or causally effective genes) are used as indicators of innate, genotypic intelligence. No one has claimed that there are "infallible measures", and Rutkowski et al. provide no citations to support their statement – they are simply asserting it.

[12] OECD (2019, p. 200), Figure I.A8.2 Self-reported effort in PISA 2018.

[13] The second conjecture variable: $r = +.21$ ($N = 79$) – if they "have invested less effort in the PISA test because it is not relevant for marks". The results here are somewhat better, but their meaning is unclear.

6 ILSA results adjustments

Rutkowski et al. (2024, p. 4) criticized "a downward score adjustment of 42 points for countries with study participants that are an average of one year older than ... the international mean age for grade-based studies" done by Rindermann in older analyses. However, they did not describe why such adjustments were made. The goal was to find population estimators of cognitive ability, not only results from youth in school. There are four problems with student assessment studies:

(1) Students are of different ages in grade level studies (TIMSS and PIRLS) – older students have an advantage, younger students have a disadvantage.

(2) Enrollment rates differ between countries. It can be assumed that children outside of school have lower cognitive ability (fewer learning opportunities plus selection effects).

(3) In some countries, only certain regions took part. Participating regions are expected to have better schools, smarter, more modernized populations, etc.

(4) In some countries, the results are very implausible, suggesting sampling errors or fraud (e.g., Kazakhstan in TIMSS 2007, Cuba in LLECE 1997 and 2005/2006).[14]

If one is interested in national cognitive ability scores to explain economic growth (which is also in the interest of an economic organization like the OECD, which runs PISA), one has to correct this. As reported by Rutkowski et al., in 2007, Rindermann corrected one year of over- or under-age with 42 SASQ points.[15] In 2018, Rindermann only applied a 14 SASQ points correction (see a longer five-page description and justification in Rindermann, 2018, Appendix, pp. 7–11). The corrections are now substantially smaller, but are the corrected values better than the uncorrected ones? Rutkowski et al. (2024, p. 4) speculated:

> *"These adjustments, with few exceptions, further penalize lower performing, Southern hemisphere countries and reward, especially, Northern and Western European countries."*

However, this is not true, as the author wrote in 2018 (Rindermann, 2018, Appendix, p. 11):

> *"Countries with the largest gains due to all corrections are (for CA total {cognitive ability}): Brunei (+6 IQ points), Belize and Tunisia (+5), Comoros and Cambodia (+3) and Korea-North (+2). The greatest downward corrections are observable for: Tajikistan and Uzbekistan (-7), Vietnam (-6), Mauritania and Gabon (-5), and Belarus (-4)."*

All adjustments serve the purpose of finding better estimates of the cognitive ability level of a *society* in order to explain, for example, prosperity, politics, and the successful management of technological modernity (cognitive human capital theory). A corrected student assessment study compared to an uncorrected one always showed slightly higher correlations with different GDP indicators (Rindermann, 2018, p. 227, Table 10.1: $r = .77$ vs. $.74$, $.78$ vs. $.76$, $.47$ vs. $.43$, $.68$ vs. $.66$, $.55$ vs. $.51$, $.68$ vs. $.66$, $.59$ vs. $.57$, $.69$ vs. $.66$.

The aim here is not to find the *absolute* truth, but to search for *relatively* more truth. We have to live with a certain degree of uncertainty in research (from a psychological point of view: one needs tolerance for ambiguity). The pattern of correlations shows that the corrections are on the right track. Critics of the corrections (a legitimate critique) are expected to propose alternative, more robust estimation methods. Of course, there are additional problems in these analyses — for example, that the individuals tested are students, not economically active adults, and that the relationship between cognitive ability and wealth is likely reciprocal, making unidirectional assumptions simplistic. Finally, the relationships can also be curvilinear.

[14] LLECE: Laboratorio Latinoamericano de Evaluacion de la Calidad de la Educacion.

[15] SASQ: Student assessment studies quotient, the scale that student assessment studies use with $M = 500$ and $SD = 100$ (Rindermann, 2018, p. 93).

7 Lynn and Vanhanen's IQ estimates based on neighbouring countries

Rutkowski et al. (2024, p. 4) criticized Lynn and Vanhanen's IQ estimates based on neighbouring countries: "such an approach is patently wrong". As above, such a claim can be tested empirically. For example, we can use the 2012 data set (Lynn & Vanhanen, 2012) and compare the measured results added here with the earlier estimated results (Lynn & Vanhanen, 2002). The correlation is r = .92 ($N = 48$ countries). The mean IQ estimated in 2002 for these 48 countries was 82.25, the measured IQ of 2012 was 81.04. The results measured later were marginally lower than Lynn's estimates. The same was done with the 2019 added measured data (Lynn & Becker, 2019) resulting in a lower but still high correlation ($r = .79$, $N = 52$ countries) and a slightly larger mean difference (2002 estimated $>$ 2019 measured: IQ = 83.31 vs. 81.06). This means that the estimation procedure worked well. Additionally, the estimates based on neighbouring countries did not underestimate the population IQs, but rather slightly overestimated them. I had always suspected that estimates derived from values of countries participating in studies overestimate the values of countries not participating in studies, but what the studies show is that my old corrections of -5 or -3 IQ points for estimated values were too high (Rindermann, 2007, p. 677; Rindermann, 2018, Appendix, p. 10). A similar analysis was carried out by Recueil (2025):

> *"The simplest way to check the robustness of Lynn's national IQs is to compare his imputed national IQs to subsequently sampled national IQs. I'll do this with his much maligned 2002 and 2012 national IQs. To assess the validity of the imputed IQs I'll correlate them with our current best national IQs {Jensen & Kirkegaard, 2024} and the World Bank HLOs {Harmonised Learning Outcomes; Angrist et al., 2021}. Lynn's 2002 imputed national IQs correlate at r = .90 with our current best national IQs and .72 with HLOs. 102 countries were imputed which we now have estimates for. ... Lynn's 2012 imputed national IQs correlate at $r = .92$ with our current best national IQs and .76 with HLOs. 66 countries were imputed which we now have estimates for."*

As before, the question remains why Rutkowski et al. did not simply calculate this themselves. Then their criticism would have been unnecessary. Is this a correct procedure in terms of the usual scientific standards?

Finally, let me emphasize that there is no such thing as error-free data at the international level. The data on GDP, for example, are sometimes bizarre (see Rindermann, 2025).[16] It cannot be a realistic requirement to use only error-free data. Instead, it is necessary to take account of errors in analyses and correct them where possible or check the robustness of results using other data and methods (Hu, 2024). The data from sub-Saharan Africa in particular are of dubious quality, are often missing, and imputations are also of little help here. However, there are some regional student achievement studies for sub-Saharan Africa that can be used (e.g., SACMEQ, MLA, PASEC).[17]

8 Accusations of "fixed" and "trend" analyses of intelligence

Rutkowski et al. (2024) accused international comparative intelligence research of advocating a concept of fixed, static intelligence (e.g., p. 2): "Idea that differences in intellectual functioning are largely genetic and that there is little room to improve learning or intelligence through, for instance, external support or

[16] For example, according to the International Monetary Fund (IMF), Ireland (as of 2023) is the richest country in the world, Guyana is wealthier than France, and Japan is poorer than Andorra. Moreover, there are large discrepancies between different sources for the same variable. In the case of Ireland, for instance, the IMF reports a per capita GDP of $145,196; the World Bank, $126,905; and the CIA, $102,500. Although the data are not from exactly the same year (ranging from 2021 to 2023), such wide variations should still be surprising. Source: https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(PPP)_per_capita (16 August 2023).

[17] SACMEQ: Southern and Eastern Africa Consortium for Monitoring Educational Quality; reading and mathematics in sixth grade. MLA: Monitoring Learning Achievement; literacy, numeracy and life skills in fourth grade. PASEC: Programme d'Analyse des Systemes Educatifs; French and math in second and fifth grade.

intervention." However, the known high heritabilities relate to inter-individual differences, not to individual development or historical developments, for example, the FLynn effect in the 20th century.[18] Patterns of inter-individual differences and patterns of country-specific differences can be quite stable and at the same time, (non-restandardized) intelligence can increase significantly with age or over time. To my knowledge, there are no researchers who exclude environmental factors (also see an analysis of educational factors for country differences; Rindermann & Ceci, 2009). One year of schooling increases cognitive ability by about 3 IQ points (Ritchie & Tucker-Drob, 2018).

Simultaneously accusing one of representing a concept of fixed intelligence and mentioning and criticizing a trend analysis of changes in intelligence over time is contradictory. However, our attempt at a statistical trend analysis (Rindermann & Becker, 2023, p. 8) explicitly stated that a purely statistical approach is not satisfactory:

> *"Predictions based solely on a statistical model, i.e. predictions based only on past developments of a few decades and extrapolating them into the more distant future, are very likely to be wrong."*

At least some theoretical assumptions (e.g., about the choice of a linear or curvilinear model, both of which were compared) are always necessary. Another problem is the usually arbitrary historical starting point of a development that is used to extrapolate future developments (depending on the availability of data; also see for the US: Rindermann & Pichelmann, 2015). Theory-based forecasting models are probably superior (e.g., Rindermann, 2018, 2023b).

Rutkowski et al. (2024, p. 5) also pointed out that the prediction uncertainty increases the further into the future the predictions extend. This is certainly a sensible suggestion that should be taken up in future predictions.

9 Framing, nonsense, politics and integrity (research ethics and behavior of researchers)

Some readers of the article by Rutkowski et al. (2024) may find the sober, fact-based response presented here unworldly and quixotic, as if the criticism by Rutkowski et al. were genuinely about science and the advancement of knowledge. Whether it actually concerns science is questionable for several reasons; at the very least, it is doubtful that this qualifies as good science. Why? Rutkowski et al. (2024) made several scientifically dubious claims without providing adequate arguments or empirical evidence:

- It denied that a test measures a construct without defining the construct.
- It denied that a test measures a psychological construct without providing analyses of content and cognitive processes.
- It claimed that there are no correlation studies and made the accusation of "ecological fallacy" (p. 4) when there are several studies at different data levels.
- It claimed that effort explains the differences in country results in ILSAs without conducting an empirical analysis, but such an analysis shows the opposite: higher effort is associated with lower PISA results (in country comparisons).
- It claimed that the result adjustments are wrong without comparatively examining their utility.
- It claimed that estimates are wrong without comparing them with measured data.

[18] Two researchers, Flynn and Lynn, or Lynn and Flynn, rediscovered the secular IQ rise, which is why we call it the "FLynn effect".

- It simultaneously criticized a concept of intelligence as fixed and a historical trend analysis of intelligence.

Added to this are blatant mistakes, such as fictitious or false first names: Rutkowski et al. (p. 3) began their chapter on "important developments in intelligence research" with the mention of an "important figure at the outset of intelligence research" "Alfred Simon"; however, this person did not exist, but a "Théodore Simon" did. Then there is also a strangely obsessive preoccupation with recapitulation theory (five times in the text), which is unknown in intelligence research.

Rutkowski et al. (p. 3) then repeat the frequently mentioned claim (e.g., Gould, 1981) that Goddard (1917) allegedly stated that "83% of Jews, 80% of Hungarians, 79% of Italians, and 87% of Russians are feeble-minded." However, Goddard never made this claim. He emphasized several times in his article that such statements should not be made – for example, in the second sentence of the summary: "2. The study makes no determination of the actual percentage, even of these groups, who are feeble-minded" (Goddard, 1917, p. 243). His sole aim was to test whether the newly developed IQ tests would also be useful for identifying mental disabilities in migrants from preselected groups. His aim is evident from the original text ("no determination of the actual percentage"), which Rutkowski et al. did not consult; Goddard is cited only via a secondary source. How can anyone demonstrate such egregious negligence and spread misinformation and hostility toward scientists?

Political framing is particularly important to Rutkowski et al. They devoted three to four pages of their six-page text to political topics and free associations that came to mind. Names like Ernst Haeckel, Charles Darwin, Jean-Baptiste Lamarck, Herbert Spencer and Henry Goddard appear. What on earth could Ernst Haeckel possibly have to do with PISA? Rutkowski et al. have spent a lot of text and space expressing, in a negative way, what they do not appreciate; examples of words that they used: "debunked science", "shoddy data", "death of millions", "holocaust", three times "inferior", seven times "racial hierarchy" and no fewer than thirteen times "eugenics". To what extent are these people and attributes relevant to answering the question of whether ILSAs measure intelligence?

They construct questionable associations and engage in skewed comparisons, which prevent them from recognizing the scientific positions of others amidst the haze of their own ideological bias. One gets the impression that they were so obsessed with political issues that they lacked both the time and the mental energy to empirically test their own theories. It is up to the reader to decide whether the article by Rutkowski et al. was more about politics and spreading hate against other researchers (see the words quoted above) than about the progress of science.

Intelligence researchers have faced threats in the past, which in extreme cases have resulted in some requiring police protection (e.g., Arthur Jensen) and others being dismissed (often overturned by the courts as unlawful, e.g., Helmut Nyborg) (Nyborg, 2003; Scarr, 1987). Since Rutkowski et al. are part of a prevailing (left-wing) zeitgeist in universities and the media, they appear less compelled to adhere to truth and intellectual integrity (Rindermann, 2023a). There is a great asymmetry of power here, intelligence researchers are in a small minority and are attacked as a minority and are hardly listened to. For members of a dominant majority, the disregard of scientific rules is tolerated within their milieu. Three further passages from the text can exemplify this:

(1) Example 1. On page 5, Rutkowski et al. (2024) wrote:

> *"Given that the authors {Rindermann & Becker, 2023} use the concept of 'genotypic intelligence' and 'genetic intelligence' (p. 205), the conclusion that entire populations are intellectually disabled on measures of intelligence is particularly disturbing, given that a deterministic view on intelligence leaves little room for these countries to improve."*

First of all, we used the term "genotypic intelligence" to describe the work of *other* authors:

> *"Despite the observable rise in intelligence test scores, other researchers assume that genotypic intelligence has been declining for some time. The 'co-occurrence model' assumes two simultaneous processes: an increase in IQ test scores that has nothing to do with the g-factor of*

> *intelligence and a decrease in genetic intelligence that would be on g ({two publications of other authors cited})." (Rindermann & Becker, 2023, p. 1).*

Rutkowski's approach here is clearly disingenuous, as they misrepresent a description of the state of research as the position of its authors. Second, our study deals with changes over time, leading to an increase of about 10 IQ points in the 21st century, from the abstract: "IQs would increase by about 10 IQ points by 2100 (international mean IQ 101)." (p. 1) And the increases are predicted to be especially large in developing countries, about 15 to 22 IQ points. A quote from our text: "The slowdown of the FLynn effect in developed countries, but its persistence in developing countries, may lead to a reduction in international disparities." (p. 8) We explicitly wrote about changes and improvements!

(2) Example 2. Rutkowski et al. (2024) start and end their article with referring to Stephen J. Gould and his book "The Mismeasure of Man". In this book, Gould alleged that researchers had cheated due to a racist motif. This was particularly ascribed to Samuel George Morton (1799–1851), an American anthropologist. With the help of craniometry, Morton's conclusion was that Europeans have, on average, larger brains than Native Americans and Africans. Gould claimed that this outcome was the result of an unconscious manipulation of the data by Morton, motivated by his "prejudices" ("finagling"). However, a student (Michael, 1988) reviewed Morton's data and found no systematic error, only a lack of precision. Michael's results were confirmed by Lewis et al. (2011), who saw systematic errors in Gould's book, but only unsystematic errors in Morton's work. These unsystematic errors actually led to lower estimates of racial group differences – so if there is still some kind of prejudice involved, it may be more of an effect of some egalitarian agenda:

> *"Clearly, Morton was not manipulating samples to depress the 'Indian' mean, and the change was trivial in any case (0.3 in^3). In fact, the more likely candidate for manipulating sample composition is Gould himself in this instance. In recalculating Morton's Native American mean, Gould reports erroneously high values for the Seminole-Muskogee and Iroquois due to mistakes in defining those samples and omits the Eastern Lenape group entirely, all of which serve to increase the Native American mean and reduce the differences between groups. ... The summary table of Morton's final 1849 catalog has multiple errors. However, had Morton not made those errors his results would have more closely matched his presumed a priori bias. Ironically, Gould's own analysis of Morton is likely the stronger example of a bias influencing results." (Lewis et al., 2011, pp. 3, 5; references omitted)*

John Michael sent his results to Stephen J. Gould but he never responded. However, considering the results of others and dealing with criticism is essential for an epistemic attitude (searching for truth) and for scientific progress (finding new truth). And not dealing with them indicates a non-epistemic attitude (pursuing goals other than truth; Rindermann, 2018).

Psychologically, it is striking that Gould accused others of bias while displaying bias himself. Such a projection is indicative of a poorly integrated cognitive system. E.g., Blinkhorn (1982, p. 506) on Gould:

> *"The theme of this {Gould's} particular book is that since science is embedded in society, one must expect to find the prejudices of the age presented by scientists as fact. Most authors, given such a theme, would be content to document and catalogue instances in support of the proposition. Gould, however, goes one better by writing a book which exemplifies its own thesis. It is a masterpiece of propaganda, researched in the service of a point of view rather than written from a fund of knowledge."*

This makes Gould far less credible than those he criticizes for lacking integrity. Other researchers, such as Russell Warne (2019), even charged Gould with explicit "lying", which Gould presumably regarded as politically legitimate:

> *"It is likely that Gould thought that his 'rhetorical strategies,' if I can call them that (. . .), were justified because of his high-minded politics. In this way, he was not unlike the pious religious*

> *fanatic who believes that inventing stories of miracles is acceptable if it strengthens the faith of others and adds more believers to the flock. Instead of 'lying for God,' though, Gould was lying for social justice."*

Gould's bias and his errors have been known for a long time and have been published by different authors at different times in different journals (e.g., the several articles mentioned above or Rushton, 1997). Those who still refer to Gould today are guided by political considerations, do not value careful work, or approach the criterion of truth with less rigor. What would Harry G. Frankfurt call it? If a contribution shows no interest in the truth and the expansion of our knowledge, that would mean that it stands for "bullshit" (Frankfurt, 1986). The reader should decide whether this also applies to Rutkowski et al.

(3) Example 3. Rutkowski et al. are particularly obsessed with the Nazi theme. Like a dealer in verbal devotional objects, they throw around terms such as "holocaust", "racial hierarchy" and "eugenics". Contrary to what Rutkowski et al. suggest, the National Socialists themselves *opposed* intelligence research — a position they share with Rutkowski et al. A striking parallel can also be seen in their mode of argumentation: Ideas they reject on ideological grounds are equated with moral evil within their worldview and consequently regarded as false. In the National Socialist view (Jaensch, 1938), intelligence measurement would be an instrument "of Jewry" to "fortify its hegemony" (p. 3), the selection in schools according to intelligence would stand for a "testing system of Jewish origin" (p. 4), especially the concept of intelligence as a "one-dimensional scale" (p. 3) (quotes translated by HR).[19] In such a climate of hate and hatred, scientific principles are no longer respected (Cofnas, 2016).

Agitation against intelligence research is nothing new, but over time, good data and scientific approaches have established many facts about intelligence, including what ILSA tests and intelligence tests have in common.

10 Concluding remarks

The work of Rutkowski et al. (2024) is riddled with glaring errors and reveals ignorance — and even folly — by failing to recognize that claims must be supported by evidence and empirical testing. Particularly egregious are Rutkowski et al.'s invention of false first names and their attribution of statements to scientists taken from secondary sources. The publication of such work constitutes a scandal for which not only the authors, but also the reviewers and editors of the journal, bear responsibility. There is no justification for publishing substandard work or for promoting hostility toward scientific reasoning and scientists. Moreover,

[19] Documentation of the difficult-to-access text, original text describing the Nazis' political-polemical position *against* intelligence research:

> *"Die Art, in der die Intelligenzprüfung und Auslese in der verklingenden Epoche vollzogen wurde, hatte den Erfolg, die Herrschaft des Gegentypus und damit auch die Vormachtstellung des Judentums innerhalb der Kulturvolker immer mehr zu befestigen." {"The way in which the testing of intelligence and selection were performed in the bygone epoch was successful in further fortifying the hegemony of the antagonistic typus, and with it also the pre-eminence of Jewry within the civilized nations."} (Jaensch, 1938, p. 3)*

> *"Aber die Intelligenzprüfung der verklungenen Epoche ruhte durchweg auf zwei fehlerhaften Voraussetzungen, die sich für unsere Volkwerdung in verhängnisvoller Weise auswirken mußten: 1. Man stellte sich die Intelligenz als eine eindimensionale Größe vor, in der es nur eine Abstufung nach 'Größer' und 'Geringer' gibt; der Physiker wurde sagen, als einen 'Skalar'." {"But the intelligence testing of the bygone epoch rested throughout on two faulty preconditions which were bound to affect our nation's development in a disastrous way: 1. One posited intelligence as a one-dimensional scale in which there exists only a gradation of "greater" or "lesser"; the physicist would say, as a 'scalar'."} (Jaensch, 1938, p. 3)*
>
> *"Unter diesen Umständen musste so gut wie zwangsläufig ein Prüfungssystem jüdischen Ursprungs zur Herrschaft gelangen, da ja das Judentum den Gegentypus, seine Normen und ‚Werte', in besonders reiner Form vertritt." {"Under these circumstances it was nearly inevitable that a testing system of Jewish origin would achieve predominance, because Jewry represents the antagonistic typus, its norms and 'values', in its purest form."} (Jaensch, 1938, p. 4)*

as in the case of Samuel Greiff, editor of *Learning and Individual Differences*, denying a response violates fundamental scientific and ethical standards.

Declaration of competing interest

The author has no competing interests.

References

Altinok, N., & Diebolt, C. (2024). Cliometrics of learning-adjusted years of schooling: Evidence from a new dataset. *Cliometrica*, *18*(3), 691–764.

Angrist, N., Djankov, S., Goldberg, P. K., & Patrinos, H. A. (2021). Measuring human capital using global learning data. *Nature*, *592*, 403–408.

Bartels, M., Rietveld, M. J. H., Baal, G. C. M. v., & Boomsma, D. I. (2002). Heritability of educational achievement in 12-year olds and the overlap with cognitive ability. *Twin Research*, *5*(6), 544–553.

Bates, T. C., & Gignac, G. E. (2022). Effort impacts IQ test scores in a minor way: A multi-study investigation with healthy adult volunteers. *Intelligence*, *92*, 101652.

Becker, D., Coyle, Th. R., Minnigh, T. L., & Rindermann, H. (2022). International differences in math and science tilts: The stability, geography, and predictive power of tilt for economic criteria. *Intelligence*, *92*, 101646.

Blinkhorn, S. (1982). What skulduggery? *Nature*, *296*, 506.

Boman, B. (2023). Is the SES and academic achievement relationship mediated by cognitive ability? Evidence from PISA 2018 using data from 77 countries. *Frontiers in Psychology*, *14*, 1045568.

Brunner, M. (2008). No g in education? *Learning and Individual Differences*, *18*, 152–165.

Burgoyne, A. P., & Engle, R. W. (2020). Attention control: A cornerstone of higher-order cognition. *Current Directions in Psychological Science*, *29*(6), 624–630.

Cattell, R. B. (1987/1971). *Intelligence: Its Structure, Growth and Action*. Amsterdam: Elsevier.

Cofnas, N. (2016). Science is not always “self-correcting”: Fact-value conflation and the study of intelligence. *Foundations of Science*, *21*(3), 477–492.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

Coleman, W., & Cureton, E. E. (1954). Intelligence and achievement. *Educational and Psychological Measurement*, *14*, 347–351.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*, 13–21.

Frankfurt, H. G. (1986). Bullshit. *Raritan Quarterly Review*, *6*(2), 81–100.

Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, *24*(1), 13–23.

Goddard, H. H. (1917). Mental tests and the immigrant. *Journal of Delinquency*, *2*, 243–277.

Gould, S. J. (1981). *The Mismeasure of Man*. New York: Norton.

Harackiewicz, J. M., Durik, A. M., Barron, K. E., Linnenbrink-Garcia, L., & Tauer, J. M. (2008). The role of achievement goals in the development of interest: Reciprocal relations between achievement goals, interest, and performance. *Journal of Educational Psychology*, *100*(1), 105–122.

Hattie, J. (2023). *Visible Learning, the Sequel: A Synthesis of over 2,100 Meta-Analyses Relating to Achievement*. London: Routledge.

Haworth, C. M. A., Kovas, Y., Dale, P. S., & Plomin, R. (2008). Science in elementary school: Generalist genes and school environments. *Intelligence*, *36*, 694–701.

Hu, M. (2024, July 3). National IQ papers must be retracted: Why Kevin Bird and Rebecca Sear don't get it. https://humanvarieties.org/2024/07/03/national-iq-papers-must-be-retracted-why-kevin-bird-and-rebecca-sear-dont-get-it

Jaensch, E. R. (1938). Grundsätze für Auslese, Intelligenzprüfung und ihre praktische Verwirklichung. [Principles for selection, intelligence measurement and their application.] *Zeitschrift für angewandte Psychologie und Charakterkunde*, *55*, 1–14.

Jensen, A. R. (2006). *Clocking the Mind: Mental Chronometry and Individual Differences*. Amsterdam: Elsevier.

Jensen, S., & Kirkegaard, E. O. W. (2024). National IQs and socioeconomic development. *PsyArXiv*. https://doi.org/10.31234/osf.io/bx86g

Jussim, L. (2020). The Orwelexicon: Neologisms for bias and dysfunctions in academia, or the DSM 666. Medium. https://medium.com/@leej12255/an-orwelexicon-for-bias-and-dysfunction-in-academia-neologisms-for-the-insufficiently-woke-a3e5bfc2953

Kaufman, S. B., Reynolds, M. R., Liu, X., Kaufman, A. S., & McGrew, K. S. (2012). Are cognitive g and academic achievement g one and the same g? An exploration on the Woodcock-Johnson and Kaufman tests. *Intelligence*, *40*(2), 123–138.

Kelley, T. L. (1927). *Interpretation of Educational Measurements*. New York: World Book Company.

Kovacs, K., & Conway, A. R. A. (2019). A unified cognitive/differential approach to human intelligence: Implications for IQ testing. *Journal of Applied Research in Memory and Cognition*, *8*(3), 255–272.

Lewis, J. E., Degusta, D., Meyer, M. R., Monge, J. M., Mann, A. E., & Holloway, R. L. (2011). The mismeasure of science: Stephen Jay Gould versus Samuel George Morton on skulls and bias. *PLoS Biology*, *9*(6), e1001071.

Lohman, D. F., & Hagen, E. P. (2002). *CogAT. Form 6*. Itasca: Riverside.

Lynn, R., & Becker, D. (2019). *The Intelligence of Nations*. London: Ulster Institute for Social Research.

Lynn, R., & Vanhanen, T. (2002). *IQ and the Wealth of Nations*. Westport: Praeger.

Lynn, R., & Vanhanen, T. (2012). *Intelligence. A Unifying Construct for the Social Sciences*. London: Ulster Institute for Social Research.

Mammadov, S. (2022). Big Five personality traits and academic performance: A meta-analysis. *Journal of Personality*, *90*(2), 222–255.

Michael, J. S. (1988). A new look at Morton's craniological research. *Current Anthropology*, *29*, 349–354.

NEPS. (n.d.). Provision of plausible values. Retrieved March 29, 2024 from www.neps-data.de/Data-Center/Overview-and-Assistance/Plausible-Values

Nyborg, H. (2003). The sociology of psychometric and bio-behavioral sciences: A case study of destructive social reductionism and collective fraud in 20th century academia. In H. Nyborg (Ed.), *The Scientific Study of General Intelligence. Tribute to Arthur R. Jensen* (pp. 441–502). Oxford: Pergamon.

OECD. (2019). *PISA 2018 results (Volume I): What students know and can do*. Paris: OECD.

OECD. (2024, March 22). What is PISA? Retrieved from www.oecd.org/pisa

Pokropek, A., Marks, G. N., & Borgonovi, F. (2022). How much do students' scores in PISA reflect general intelligence and how much do they reflect specific abilities? *Journal of Educational Psychology*, *114*(5), 1121–1135.

Recueil, C. (2025, January 16). National IQs are valid. National IQ estimates are robust, reliable, and realistic. https://www.cremieux.xyz/p/national-iqs-are-valid

Rindermann, H. (2007). The *g*-factor of international cognitive ability comparisons: The homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests across nations. *European Journal of Personality*, *21*, 667–706.

Rindermann, H. (2018). *Cognitive Capitalism: Human Capital and the Wellbeing of Nations*. Cambridge: Cambridge Univ. Press. Appendix: https://figshare.com/s/77ec7070b96005ab314f

Rindermann, H. (2023a). The advantages of having a minority viewpoint in politicized psychology: A case study of intelligence research. In C. L. Frisby, R. Redding, W. T. O'Donohue, & S. O. Lilienfeld (Eds.), *Ideological and Political Bias in Psychology* (pp. 709–741). Cham: Springer.

Rindermann, H. (2023b). The future of intelligence in Germany: Assumptions, models and predictions. In R. Lynn (Ed.), *Intelligence, Race and Sex: A Tribute to Helmuth Nyborg at 85* (pp. 224–265). London: Arktos.

Rindermann, H. (2024). Why are there differences across German states in student achievement and cognitive ability? *Heliyon*, e25043.

Rindermann, H. (2025). Low cognitive ability estimates in developing countries: A statistical analysis of their credibility. *Human Evolution*, *40*(3–4), 257–284.

Rindermann, H., & Baumeister, A. E. E. (2015). Validating the interpretations of PISA and TIMSS tasks: A rating study. *International Journal of Testing*, *15*(1), 1–22.

Rindermann, H., & Becker, D. (2023). The future of intelligence: A prediction of the FLynn effect based on past student assessment studies until the year 2100. *Personality and Individual Differences*, *206*, 112110.

Rindermann, H., & Ceci, S. J. (2009). Educational policy and country outcomes in international cognitive competence studies. *Perspectives on Psychological Science*, *4*(6), 551–577.

Rindermann, H., & Pichelmann, S. (2015). Future cognitive ability: US IQ prediction until 2060 based on NAEP. *PLoS ONE*, *10*(10), e0138412.

Rindermann, H., Flores-Mendoza, C., & Mansur-Alves, M. (2010). Reciprocal effects between fluid and crystallized intelligence and their dependence on parents' socioeconomic status and education. *Learning and Individual Differences*, *20*, 544–548.

Ritchie, S. J., & Tucker-Drob, E. M. (2018). How much does education improve intelligence? A meta-analysis. *Psychological Science*, *29*, 1358–1369.

Rushton, J. Ph. (1997). Race, intelligence, and the brain: The errors and omissions of the revised edition of S. J. Gould's *The Mismeasure of Man* (1996). *Personality and Individual Differences*, *23*, 169–180.

Rutkowski, L., Rutkowski, D., & Thompson, G. (2024). What are we measuring in international assessments? Learning? Probably. Intelligence? Not likely. *Learning and Individual Differences*, *110*, 102421.

Scarr, S. (1987). Three cheers for behavior genetics: Winning the war and losing our identity. *Behavior Genetics*, *17*(3), 219–228.

Scharl, A., Carstensen, C. H., & Gnambs, T. (2020). Estimating plausible values with NEPS data: An example using reading competence in starting cohort 6. Bamberg: NEPS Survey Paper No. 71.

Schneider, W., & Niklas, F. (2017). Intelligence and verbal short-term memory/working memory: Their interrelationships from childhood to young adulthood and their impact on academic achievement. *Journal of Intelligence*, *5*(2), 26.

Shakeshaft, N. G., Trzaskowski, M., McMillan, A., Rimfeld, K., Krapohl, E., Haworth, C. M. A., Dale, P. S., & Plomin, R. (2013). Strong genetic influence on a UK nationwide test of educational achievement at the end of compulsory education at age 16. *PLoS ONE*, *8*(12), e80341.

Stanek, K. C., & Ones, D. S. (2023). Meta-analytic relations between personality and cognitive ability. *Proceedings of the National Academy of Sciences*, *120*(23), e2212794120.

Steinberg, L. (1996). *Beyond the Classroom*. New York: Simon & Schuster.

Vu, T. V., Scharmer, A. L., van Triest, E., van Atteveldt, N., & Meeter, M. (2024). The reciprocity between various motivation constructs and academic achievement: A systematic review and multilevel meta-analysis of longitudinal studies. *Educational Psychology*, *44*(2), 136–170.

Warne, R. T. (2019, March 19). The mismeasurements of Stephen Jay Gould. Quillette. https://quillette.com/2019/03/19/the-mismeasurements-of-stephen-jay-gould

Warne, R. T. (2023). National mean IQ estimates: Validity, data quality, and recommendations. *Evolutionary Psychological Science*, *9*, 197–223.

Zaboski, B. A., Kranzler, J. H., & Gage, N. A. (2018). Meta-analysis of the relationship between academic achievement and broad abilities of the Cattell-Horn-Carroll theory. *Journal of School Psychology*, *71*, 42–56.