Intelligence 98 (2023) 101756

Contents lists available at ScienceDirect

Intelligence

journal homepage: www.elsevier.com/locate/intell

General intelligence in middle school students from different Russian regions: Results of PISA-like tests

Nikita Kolachev[a,*], Galina Kovaleva[b]

[a] International Laboratory of Positive Psychology of Personality and Motivation, HSE University, 4 building 2 Slavyanskaya ploshchad', Moscow, Russian Federation
[b] Center for Education Quality Assessment, Institute for Strategy of Education Development of the Russian Academy of Education, 16 Zhukovskii str., Moscow, Russian Federation

ARTICLE INFO

Keywords:
Intelligence
G-factor
General cognitive ability
Bifactor model
PISA

ABSTRACT

This study is aimed at investigating the contribution of the general intelligence factor if six PISA domains (reading, mathematical, scientific, financial literacies, global competence, and creative thinking) are combined in one measurement instrument. For achieving our goal, items based on the PISA frameworks are developed, students in grades 5–8 from three different Russian regions are assessed, and three IRT models (unidimensional, multidimensional, and bifactor) are applied to process the data. In addition, the correlations from the multidimensional model are estimated to examine the degree of cognitive specificity and mixture modeling is implemented to investigate ability differentiation across grades. Statistical analysis reveals that the bifactor model comprising one general and six specific factors, has a better fit in each grade. Based on this model, we compute the variance explained by the general factor, with the estimates varying between 60% and 70%. In general, the pure variance explained by specific factors does not exceed 10%. The correlations are above 0.40 in each grade and the averaged associations tend to increase from 6th to 8th grade, although they are smaller in years 6 and 7 compared to year 5. The general ability differentiation effect is observed in grades 6 to 8 and is not present in grade 5. Specific ability differentiation is more pronounced in reading literacy, especially in grade 5 to 7. The results obtained are discussed from the perspective of the ability and developmental differentiation/dedifferentiation problem.

1. Introduction

Several studies (Brunner, 2008; Pokropek, Marks, & Borgonovi, 2022; Pokropek, Marks, Borgonovi, Koc, & Greiff, 2022) have demonstrated that the results of the Programme for International Student Assessment (PISA) are better described with a bifactor model that includes one general factor andseveral specific domains. On Polish data it was found that the general factor dominates over domain-specific ones and explains almost 80% of variance in students' responses (Pokropek, Marks, & Borgonovi, 2022). How would such results be explained? Cromley (2009) suggests two potential interpretations. First, much variation in item performance might be determined by reading literacy: those who have higher levels of reading comprehension demonstrate higher results in other domains. Second, there is another latent factor (some domain-general factor) that causes the results in the PISA disciplines. Pokropek, Marks, Borgonovi, Koc, and Greiff (2022) on a sample of 33 OECD countries revealed that the reading-factor does not explain much variation in students' responses compared to the general one, called intelligence.

Why may we speak about modeling intelligence with PISA or PISA-like measurement instruments? The first reason is operationalization of the PISA domains (literacies) that highly overlaps with some definitions of intelligence. As PISA organizers underscore, it measures “students' capacity to apply their knowledge and skills in key areas, and to analyse, reason and communicate effectively as they identify, interpret and solve problems in a variety of situations” (OECD, 2019c, p. 26). So, in reliance on the mentioned conceptualization the main characteristics of literacy are the following:

- Use of acquired knowledge, competencies, and skills;
- Solving a variety of problems;
- Manifestation in diverse contexts and situations of human live.

These properties intersect with the definition of intelligence used by

* Corresponding author.
E-mail address: nkolachev@hse.ru (N. Kolachev).

https://doi.org/10.1016/j.intell.2023.101756
Received 25 August 2022; Received in revised form 3 April 2023; Accepted 3 April 2023

Pokropek and his colleagues based on the operationalization of Gottfredson (1997) and Neisser et al. (1996) which is following: intelligence is "the ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, to overcome obstacles by taking thought" (Pokropek, Marks, & Borgonovi, 2022, p. 1121). The mentioned interpretation of intelligence also includes problem solving, knowledge, competencies, and skills acquisition (through learning from experience), and a variety of contexts. L. Gottfredson emphasizes that traditionally literacy tests mimic intelligence tests and functional literacy represents "a general capacity to learn, reason, and solve problems" (Gottfredson, 2005, p. 177).

Second, for the last few decades, one of the important skills has been analogical (knowledge) transfer (Willingham, 2021). It means applying old knowledge to solve new problems. This cognitive skill can be considered high-level because it requires a lot of developmental work and makes a person more flexible in problem solving (Holyoak & Koh, 1987). As a cognitive process, transfer includes three stages: retrieving from memory of an analogous situation; comparison of representations of the known and new task; evaluation of the adequacy of the solution for the new task (Gentner & Maravilla, 2017). By definition, PISA assesses students' capacity to apply knowledge and skills which reflects the process of transfer. Barnett and Ceci (2002) underline that transfer depends on the level of intelligence. Moreover, Sternberg (1985) in his work on the triarchic theory, emphasized that transfer is a part of intelligence which is related to the retrieval and performance components. Third, the expert community agrees that tests measuring literacies require more general cognitive ability or intelligence than academic achievement tests (such as TIMSS and PIRLS). For instance, Rindermann (2007, p. 692) based on a review of theoretical and empirical studies writes: "The 'literacy' concept could be included in the historically older intelligence concept". In addition, Rindermann & Baumeister (2015) point out that, according to teachers' and psychology students' opinions, PISA-type tasks are more likely to require reasoning ability and general intelligence compared to TIMSS items. Finally, empirical research confirms that results of the international large-scale assessments are highly correlated with different intelligence measures (Rindermann, 2007). Furthermore, inclusion of intelligence scores along with literacy tests results into latent variable modeling does not worsen the fit statistics and indicates unidimensionality (Pokropek, Marks, & Borgonovi, 2022). Thus, models with g-factor usually fit the data better than multidimensional ones.

The considerable presence of a general factor is described by the fact that the PISA subjects, such as reading, mathematical, and scientific literacy include a similar set of competencies (e.g., identification of information needed, understanding, and evaluation), which overlaps intelligence features (Pokropek, Marks, & Borgonovi, 2022). According to the latest PISA cycles (OECD, 2019c) reading literacy is operationalized as "understanding, using, evaluating, reflecting on and engaging with texts in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society" (p. 34); mathematical literacy is defined as "students' capacity to formulate, use and interpret mathematics in a variety of contexts" (p. 104); scientific literacy is "students' ability to engage with science-related issues and with the ideas of science, as reflective citizens" (p. 112). Empirical studies show that across many cycles and different countries, reading, mathematics, and science results are highly interrelated (Cromley, 2009; Sjøberg & Jenkins, 2020). The correlations rarely drop to <0.70 (OECD, 2019c). Recently, Pokropek, Marks, and Borgonovi (2022) on the Polish PISA-2009 sample have found that reading, mathematics, and science strongly correlate with measures of intelligence obtained with Raven's Matrices: the latent correlations varied between 0.69 (science with intelligence) and 0.89 (mathematics with intelligence) the year of the PISA study (2009) and between 0.69 (reading with intelligence) and 0.90 (mathematics with intelligence) a year later (2010). Using different latent variable models, they have demonstrated that the results are better fitted with a bifactor model where g-factor dominates at the level of 73% of the explained variance compared to the multidimensional and higher order factor models. Model fit and the g-factor prevalence do not change noticeably even if the dimension of the Raven's Matrices is excluded.

Baumert, Lüdtke, Trautwein, & Brunner (2009) underscore that the main feature of the "true" intelligence measure is its context and content independence. They posit that intelligence instruments do not involve the content-specific knowledge as opposed to achievement tests. As we posited above, the PISA assessments are distinguished by the need to use the transfer of knowledge and skills to unfamiliar situations (non--curriculum problems). These latter may involve personal, social, professional, scientific, or other contexts. Moreover, PISA conceptual frameworks include similar or the same sets of cognitive processes (competencies) across different literacies. These cognitive processes are universal. So, a student can demonstrate an ability to analyze the information in reading, mathematics, or science tasks and this still will be an analysis of information. Thus, the same competencies can be assessed in various contexts and across various literacies which makes PISA less like typical achievement tests. Moreover, the high correlation between the Raven test (content-independent) and results in mathematics and science literacy (content-dependent domains) revealed by Pokropek, Marks, and Borgonovi (2022) suggests that content independence is not necessarily a property of true measures of intelligence. In addition, the researchers write: "People with high cognitive abilities in one domain (e.g., verbal) tend also to perform better in other domains (e.g., spatial and mathematical)" (Ibid., p. 166). The high correlations across PISA domains support this idea and might demonstrate that universal cognitive processes (e.g., analysis, evaluation, application) assessed in the study might be context-independent.

We agree that students use their subject knowledge when performing PISA tasks. Nevertheless, in well-known intelligence tests, such as the Wechsler test (Wechsler, 2014), the part aimed at measuring visuospatial abilities can be provided by the experience of learning geometry; picture arrangement and completion also might be based on mathematics skills because mathematics lessons develop the ability to find relations and dependencies between objects. Finally, this test includes arithmetic tasks which are not content-free. Verbal comprehension (another domain of the Wechsler test) can be closely related to reading literacy and its development (Woolley, 2010). Furthermore, pure intelligence test scores might be prone to change under the influence of education (Brinch & Galloway, 2012). Also, researchers claim that intelligence phenotypically and genetically associates with years of education (Plomin & Von Stumm, 2018; Rietveld et al., 2014). Behavioral genetic studies suggest that intelligence is, on average, half heritable and half environmentally influenced (Plomin & Petrill, 1997); however, the effect of heritability usually increases with age (Plomin, Fulker, Corley, & DeFries, 1997). Academic achievements follow a similar pattern (Petrill & Wilkerson, 2000). In turn, there is evidence to show that heritability explains more than half of the variation in standardized achievement tests (Pokropek & Sikora, 2015).

Potentially, another fundamental mechanism underlying relations across PISA domains are executive functions – cognitive processes that help regulate and control behavior (in a broad sense). They are provided by the frontal lobes of the brain (Friedman et al., 2006). Their core components are inhibitory control, working memory, and mental flexibility (Diamond, 2013). It might be evident that every testing of cognition or academic achievements involves such mental processes as perception, attention, and memory. Inhibition can be involved in keeping attention when performing different tasks or dealing with temptations to give up. As Diamond (2013) underlines, working memory is essential for making sense of written or spoken language, mentally reordering items, translating instructions into action plans, incorporating new information into thinking, considering alternatives, and mentally deriving a general principle or seeing relations between items or ideas. Cognitive flexibility might be important in situations when one way of solving a problem is not working and an individual should come up with a new one, that is to appropriately adapt behavioral strategies (Miles et al., 2021). Despite the importance of all executive functions for cognitive testing, researchers show that, amongst them, only working memory strongly relates to crystallized and fluid intelligence (Friedman et al., 2006). Still, scholars agree that intelligence tests do not fully capture executive function (Ardila, 2018). Furthermore, a recent behavioral genetic study has demonstrated that executive functions and intelligence are more distinct in older adolescents and adults than in children and young adolescents (Gustavson et al., 2022).

1.1. Ability and developmental differentiation-dedifferentiation

A problem of ability and developmental differentiation-dedifferentiation relates to the extraction of the common intelligence factor. It is believed that the topic was first raised by C. Spearman (Spearman, 1927; Tucker-Drob, 2009). He primarily spoke about ability differentiation by showing that in normo-typical children the mean correlation across several cognitive tasks was weaker than in children with special educational needs ($r_1 = 0.47$ and $r_2 = 0.78$ respectively). This means that the general intelligence factor is more important for low-ability students compared to high-ability ones. This finding was replicated several times and has been considered empirically supported (Blum & Holling, 2017). Recent large-scale studies using contemporary sophisticated statistical methods have shown that ability differentiation significantly depends on age and maturation processes (Breit, Brunner, & Preckel, 2020): younger adolescents (10–12-year-olds) demonstrate a more pronounced effect of ability dedifferentiation, while older adolescents experience the process of ability differentiation. Moreover, in the latest systematic review Breit, Brunner, Molenaar, and Preckel (2022) have demonstrated that ability differentiation in children and adolescents is found in most of the scientific articles they analyzed.

Another important effect is age (or nowadays developmental) differentiation. It was noticed by H. Garrett (Garrett, 1938; Tucker-Drob, 2009) and was expressed in the fact that with child development their cognitive abilities become more specific, i.e., the role of the general factor decreases. There are two perspectives on this process. On the one hand, investment theory (Cattell, 1987) states that people are born with a strongly pronounced single cognitive ability, but in the process of life activity under the influence of various factors, including non-cognitive, form differentiated intellectual abilities. On the other hand, dynamic mutualism theory (Van Der Maas et al., 2006) emphasizes that, from birth, people are equipped with orthogonal cognitive abilities, that under the process of maturation and learning begin to interact closely, which eventually leads to the formation of g-factor. The existing empirical studies tend to support the developmental dedifferentiation hypothesis. For instance, Blum and Holling (2017) in a meta-analysis found a slight developmental dedifferentiation trend across 394 research papers. In the previously mentioned study conducted by Breit et al. (2020) it was clearly observed that adolescents show the developmental dedifferentiation effect with age. Also, Tucker-Drob (2009) exhibited that dedifferentiation tendency persists during childhood and adolescence. In a recent systematic review, it has been underlined that 57.2% of the effect sizes supported age dedifferentiation in children and adolescents, while 40.4% of them indicated no effect (Breit et al., 2022).

In summary, we can say that the age differentiation hypothesis has more contradictions than the ability one. Contrary to the developmental differentiation hypothesis, there is more evidence for developmental dedifferentiation. Moreover, the existing studies have tested these effects utilizing the well-known intelligence tests, but there are few studies examining these effects based on PISA-like instruments, which, as we have tried to show above, end up measuring intelligence. Investigating the contribution of general intelligence, as well as testing hypotheses about ability and developmental differentiation, on PISA-like items is important because these tasks are used in school classes and important policy decisions might be based on the results of their administration. For example, our country has established a federal program in reliance on PISA-like tests to improve students' competencies.

Considering all the importance of this topic, we suggest some further steps in it. In addition to the basic three literacies, PISA also assesses other domains, such as financial literacy, global competence, and creative thinking (OECD, 2019a; OECD, 2019b). The reason for the inclusion of these subjects into research models is that financial literacy and global competence have demonstrated strong correlations with the core disciplines (OECD, 2020). More important, they include similar cognitive processes as the core subjects: analysis, evaluation, and application (OECD, 2019b). The situation with creative thinking is less clear because we do not have PISA data as the testing procedure has taken place in 2022. Nevertheless, the problem of the association between intelligence and creativity has long occupied researchers (Jauk, Benedek, Dunst, & Neubauer, 2013), and empirical studies demonstrate that there are positive links between these constructs, although not strong (Kim, 2005). Next, we present a literature review of relations of these three additional areas of assessment with intelligence.

1.2. Financial literacy and intelligence

Financial literacy is considered an essential skill that helps achieve economic welfare of individuals (Lusardi & Mitchell, 2014). Poor financial literacy does not allow citizens to effectively plan their budget, make decisions in the field of personal or family finances (Lusardi, Mitchell, & Curto, 2010).

Financial literacy was added to the PISA assessment procedure in 2012 (OECD, 2012) and is defined as "the knowledge and understanding of financial concepts and risks, and the skills, motivation, and confidence to apply such knowledge and understanding to make effective decisions across a range of financial contexts, to improve the financial well-being of individuals and society, and to enable participation in economic life" (OECD, 2019b, p. 128). Amongst cognitive processes needed for financial literacy, PISA experts identify four: identification of financial information, analysis of information in a financial context, evaluation of information, and application of financial knowledge and understanding (Ibid.).

Finke, Howe, and Huston (2017) in their study dedicated to the effects of aging on financial literacy and cognitive abilities, revealed that a decline in fluid and crystallized intelligence was related to a decrease in financial literacy scores. D'Acunto, Hoang, Paloviita, and Weber (2019) using IQ scores, retrieved from the cognitive tests of the Finnish Defence Forces, found that IQ measures were positively correlated with financial knowledge, represented as precision of inflation expectations, and financial decisions.

Several studies demonstrated that various cognitive abilities (not necessarily measures of intelligence) were associated with financial literacy. For instance, Muñoz-Murillo, Álvarez-Franco, and Restrepo-Tobón (2020) in experimental research showed that controlling for students' socio-economic status cognitive abilities, assessed with Cognition Reflection Test, significantly predicted results of a financial literacy test. Lin and Bates (2022), in two different samples, found moderate correlations between cognitive ability measured with verbal reasoning, letter and number series, and matrix reasoning tests, and financial literacy composed of such subfactors as financial knowledge, financial competence, and time preference. Also, they concluded that financial literacy was not linked to specific cognitive components but was determined by general cognitive ability.

The latest international OECD report communicates that in all participating countries financial literacy was strongly associated with reading and mathematics literacies. The average correlations for all OECD countries are 0.83 and 0.87 respectively. Therefore, we may expect a close connection of financial literacy with g-factor.

1.3. Global competence and intelligence

In 2018, global competence became a part of the PISA assessment procedure (OECD, 2019b). It is defined as the competence to "examine local, global and intercultural issues, understand and appreciate different perspectives and worldviews, interact successfully and respectfully with others, and take responsible action toward sustainability and collective well-being" (Ibid., p. 166). The conceptual framework describes global competence as consisting of four dimensions: knowledge, skills, attitudes, and values. Also, it specifies two approaches for measuring it: objective test and self-reports. Knowledge and skills are assessed with either a cognitive test or self-reports, while attitudes only with self-reports. Values are out of the PISA scope. In our analysis, we focus on the global competence cognitive test. Its items assess the following set of cognitive processes: evaluate information, formulate arguments and explain complex situations and problems, identify and analyze multiple perspectives and worldviews, understand differences in communication, and evaluate actions and consequences. Researchers underscore that although there are many instruments intended to measure global competence, most of them include cognitive or meta-cognitive components (Liu, Yin, & Wu, 2020). It seems natural because global competence is supposed to relate somehow to knowledge, analysis, evaluation of global issues, and means of intercultural communications.

We could not find any publications directly linking global competence with intelligence. However, a part of global competence, intercultural competence, is believed to be related to emotional intelligence (Guntersdorfer & Golubeva, 2018). At the same time, recent studies support the idea that emotional intelligence is a subfactor of general cognitive ability (Evans, Hughes, & Steptoe-Warren, 2020). Moreover, according to the PISA international report (OECD, 2020), global competence is highly correlated with the core subjects ($r_{OECD} = 0.84$ with reading, $r_{OECD} = 0.73$ with mathematics, and $r_{OECD} = 0.79$ with science). The core PISA assessment areas are better described by g-factor, so the strong association of global competence with the latter might be expected.

1.4. Creative thinking and intelligence

The cycle of 2021 comprised creative thinking (OECD, 2019a) and operationalized it as "the competence to engage productively in the generation, evaluation and improvement of ideas, that can result in original and effective solutions, advances in knowledge and impactful expressions of imagination" (Ibid., p. 8). As it is clear from the definition, the PISA assessment framework of creativity relies on ideas generation and the correction of existing ones, which is perceived as divergent thinking (Benedek, Jauk, Sommer, Arendasy, & Neubauer, 2014; Kaufman, Plucker, & Baer, 2008). Divergent thinking is a part of Guilford's model of intelligence (Guilford, 1967). He believed that creativity was a subordinate construct of intelligence (Jauk et al., 2013). Some other researchers explain positive associations between intelligence and creativity by the existence of a common cognitive basis which is working memory (Benedek, Jauk, et al., 2014). When performing a creativity task, working memory is utilized, especially in generation of a new idea (Benedek et al., 2014). At the same time, intelligence measures and the working memory span are highly connected (Shelton, Elliott, Hill, Calamia, & Gouvier, 2009).

1.5. Present study

The aim of this study is threefold. First, we wish to replicate the results of Pokropek, Marks, and Borgonovi (2022) on samples of Russian middle school students. It should be emphasized that their results were obtained in a sample of the PISA study sample (9th–10th grades). The only research conducted on other grades but 9th–10th was by Saß, Kampa, and Köller (2017). However, this study was limited to examining mathematical tests. So, we are not sure whether the factor structure and the effect of the general ability would manifest in other grades (from 5th to 8th). This is important for the educational process, as the development of cognitive abilities occurs throughout schooling, not by the end of middle school when PISA conducts assessments. Moreover, researchers underscore that in cognitive tests variation the effect of g-factor increases with age (Tucker-Drob, 2009; Whitley et al., 2016).

Second, we would like to elucidate the role of general cognitive ability in a broader range of assessment areas by adding other PISA domains to the models – financial literacy, global competence, and creative thinking. This seems important to us because the development of intelligence occurs in different school subjects, and if performance on tests from different areas shows a large contribution of intelligence, this may indicate a need for integration of efforts on the part of teachers at different school subjects. To reach this goal, our test developers and we elaborated items based on the PISA frameworks. We designed specific tests on a computer-based platform that included not all but four out of six assessment areas and followed the incomplete test paradigm (see the Measures section for more details).

Third, we are encouraged to explore ability differentiation in every mentioned grade. We propose to investigate it using a model-based approach which is perceived as more reliable because it is not based on arbitrary cut-offs for dividing students into different ability groups (Breit et al., 2020). Moreover, model-based techniques help to account for the non-linearity of the test results (Reynolds, Keith, & Beretvas, 2010). Besides, we propose to consider ability differentiation not only in terms of the contribution of g-factor, but also in terms of the contribution of specific factors in different ability groups. Such analysis may help to partially explain the possible effect of ability differentiation (Feraco & Cona, 2022).

Our hypotheses are following:

H$_1$. *g-factor will explain most of the variance of students' responses;*

H$_2$. *the dominance of the general factor will increase from 5th to 8th grades;*

H$_3$. *in lower-ability students g-factor will explain more variance compared to higher-ability students in each sampled grade.*

The present study is a part of the federal project "Monitoring the formation of students' functional literacy" initiated by the Ministry of Education of the Russian Federation (http://skiv.instrao.ru/content/board1/). The main project executor is the Institute for Strategy of Education Development of the Russian Academy of Education.

2. Method

2.1. Participants

The studied samples include 26,688 fifth graders from 488 schools, 21,973 sixth graders from 627 schools, 11,067 seventh graders from 162 schools, and 38,970 of eighth graders from 957 schools. The mean age of the fifth graders is 11.94 (SD = 0.43 years), mean age of the sixth graders is 12.92 (SD = 0.42 years), age of the seventh graders is 13.94 (SD = 0.43 years), and age of the eighth graders is 14.92 (SD = 0.43 years). All samples are gender-balanced.

The participants come from three different Russian regions, pupils from 5th and 8th grade are from the same one. The regions are comparable in terms of students' academic attainments. The mentioned number of respondents make up the general population of the students of their respective regions. However, bearing in mind the large sample sizes and the modest computational abilities of our hardware, we randomly sampled 5000 participants from every named group. This number of students is close to the number that appears in the PISA sample. In addition, this number of respondents is not too computationally demanding. So, the random samples are expected to be representative since we draw them in conditions when each person of the study population has an equal probability to be in the sample (simple random sampling). The random sampling was conducted using the "sample_n" function of the dplyr package (Wickham, François, Henry, &

Müller, 2022) for R 4.0.2 (R Core Team, 2020).

To make sure that the sampled data represent the results of the full data, we examined descriptive statistics of the distributions of the ability variable (in logits) based on the unidimensional 1-parameter logistic model (Rasch, 1960). In addition, we computed the Kolmogorov-Smirnov Distance (D) which is a metric of equality of two empirical distributions; it varies between 0 (the same distributions) and 1 (different distributions) (Kolmogorov, 1933). Results presented in Table 1 demonstrate that all subsamples are almost equal to the original data.

2.2. Measures

Six test variants consisting of 4 out of 6 assessed disciplines were created. The distribution of the disciplines across the test variants is presented in Table 2. We used the block design. Every block comprised one or two textual and/or picture stimuli and items to them called testlets. The variants were equated with common blocks within grades, but not across them. So, the test items are different in every grade. The equating procedure is necessary to make sure that students' results are placed on the same interval scale and are comparable. To fill in the test variants, first we randomly assigned the blocks in every table cell. Second, if the same block was in the same position in different variants, we manually swapped blocks to account for the positional effect. All test blocks repeated twice (see Table 2). Therefore, each block was performed by one-third of the examinees.

The items were calibrated using the classical Rasch model for dichotomous items (Rasch, 1960) and Masters partial credit model for polytomous ones (Masters, 1982). These models place students' ability measures and item difficulty parameters on the same interval logit scale. The item difficulty and the student ability are linked by a logistic function which is probabilistic by its nature. Given the probabilistic feature of the models, students do not need to be administered with the whole test battery. If there are linking items across test variants, the probabilistic models can create a scale on which every item and every student will be located.

We used the concurrent linking approach which is based on simultaneous estimation of the model parameters in reliance on incomplete data matrix. The items were analyzed regarding their fit to the models. We performed the analysis based on weighted and unweighted mean-square fit values. According to Wright, Linacre, Gustafson, and Martin-Löf (1994), only those items that demonstrate fit values higher than 2 distort or degrade the measurement results. All our test items showed an adequate fit to the Rasch-family models mentioned above. In grade 5, the fit statistics varied between 0.48 and 1.65; in grade 6, between 0.33 and 2.02; in grade 7, between 0.36 and 1.95; in grade 8, between 0.29 and 2.06. We also conducted the Differential Item Functioning analysis (DIF-analysis) to compare item difficulties across different test variants (to assess the position effect). It is based on a comparison of the difficulty of the same items (in logits). A threshold for the presence of DIF is >0.64 logits in absolute magnitude; at the same time, Welch's *t*-test should be significant at $p < .05$ (Boone, Staver, & Yale, 2013). In grades 5 and 8, none of the items showed DIF. In grade 7, only one item demonstrated unequal functioning (the difference in difficulty is 0.70). In grade 6, two reading literacy items exceeded the threshold value (the differences in difficulty are 0.72 and 0.66). Despite the presence of DIF, we did not exclude these items from further analysis, because in the total set of items one or two of them would not significantly affect the result.

Items and testlets for the instrument were developed based on the PISA conceptual frameworks. The latter are universal for all grades. Nevertheless, the test materials were adapted to each grade considering students' social experience and level of education. For example, students in grades 5–6 were not offered tasks related to occupational contexts, as this is usually irrelevant for these ages in our country.

Test blocks included either dichotomous or polytomous items. Table 3 contains information about the number of items and maximum score of every test block. The test for the 5th-graders contains 84 items, for the 6th-graders – 93 items, for the 7th-graders – 92 items while the test for the 8th-graders, 85 items. Response formats comprised simple multiple-choice, complex multiple-choice, and open-ended answers.

The open-access bank of the test items is presented here (in Russian): http://skiv.instrao.ru/bank-zadaniy/.

2.3. Procedure

The data were collected in 2021 when students all over the country returned from distant to face-to-face learning. Students' parents gave their informed consent to participate in the study at the beginning of the school year. The testing procedure took place during two school lessons (40 min each) with a 10-min break between them. Students performed the tests on an online-platform in class-settings. We limited the test run time to 20 min per block. After 20 min expired, students were automatically redirected to another block or a break. The reasons for the time restrictions are as follows: first, in order not to disrupt the usual rhythm of the schools, we could only conduct assessments during two lessons (40 min + 40 min). Second, to allow students to perform items in the maximum amount of assessment areas, we decided to limit the completion of one block to 20 min. Third, we followed the block test design of the Trends in Mathematics and Science Study (TIMSS). This study provides students with four blocks of items during two school lessons, each block takes 18 (fourth grade) to 22.5 min (eighth grade) (Mullis & Martin, 2017).

When the testing procedure was finished, teachers at the respective schools assessed open-ended questions using our criteria; they were trained to do that by our subject matter specialists. In case of discrepancies or problems, our experts helped teachers to solve them.

2.4. Statistical analyses

Before testing the stated hypotheses, we performed some data processing. Students who had missing values on all test items were excluded from the data matrix. Also, those students who had medically verified cognitive abilities problems were also removed from the dataset. During the data collection process, we included two types of missing values – omitted (coded as 9) and not reached (coded as 99). “Omitted” means that a student got to an item, but for various reasons did not perform it

Table 1
Descriptive statistics of the students' ability distribution and the Kolmogorov-Smirnov Distance metric of the full samples and their random subsamples.

Grade	Sample	Min	25th percentile	M (SD)	75th percentile	Max	Kolmogorov-Smirnov Distance
5	Full	−5.57	−1.90	−1.40 (0.85)	−0.81	1.39	0.01
	Subsample	−5.57	−1.90	−1.40 (0.85)	−0.81	0.95	
6	Full	−6.62	−1.68	−1.22 (0.77)	−0.66	1.01	0.02
	Subsample	−6.62	−1.69	−1.22 (0.77)	−0.66	0.75	
7	Full	−6.59	−1.72	−1.29 (0.83)	−0.73	0.80	0.01
	Subsample	−6.32	−1.74	−1.29 (0.84)	−0.75	0.80	
8	Full	−6.56	−1.75	−1.24 (0.83)	−0.65	1.35	0.01
	Subsample	−5.31	−1.76	−1.24 (0.83)	−0.65	1.10	

Table 2
Distribution of blocks across test variants in 5th–8th -grades.

Variant	First lesson			Second lesson	
	20 min	20 min		20 min	20 min
1	Reading literacy, block 2	Creative thinking, block 2	10-min break	Science literacy, block 1	Mathematics literacy, block 1
2	Mathematics literacy, block 2	Science literacy, block 2		Financial literacy, block 1	Global competence, block 2
3	Creative thinking, block 1	Reading literacy, block 1		Global competence, block 1	Financial literacy, block 2
4	Science literacy, block 2	Global competence, block 1		Mathematics literacy, block 1	Creative thinking, block 1
5	Global competence, block 2	Financial literacy, block 1		Reading literacy, block 1	Science literacy, block 1
6	Financial literacy, block 2	Mathematics literacy, block 2		Creative thinking, block 2	Reading literacy, block 2

Table 3
Number of items and maximum score per test block.

Variable	n of items (max score)			
	5th grade	6th grade	7th grade	8th grade
RL	10 (14)	11 (17)	11 (17)	7/8 (9)
ML	6 (11)	8 (13)	8 (13)	5 (8)
SL	9 (13)	10 (13)	10 (13)	9/10 (13)
FL	8 (12)	8 (12)	8 (12)	10 (16)
GC	5 (8)	5 (8)	5 (8)	6 (9)
CT	3/5 (9)	4/5 (7)	4 (7)	4/5 (9)

Note. RL – reading literacy, ML – mathematics literacy, SL – scientific literacy, FL – financial literacy, GC – global competence, CT – creative thinking.

and moved to the next one. “Not reached” means that a student did not have time to perform an item due to time constraints. In the database, we recoded 9 to 0 because a student had the opportunity to complete the task but did not do so; and 99 was recoded to the missing value, that is, lack of information. We did not remove any outliers because they represent real variations in the population.

To investigate the best factor structure, we implemented three different models: unidimensional, multidimensional (with 6 correlated factors), and bifactor. The dimensionality analysis is based on the IRT 2-parameter logistic model (Birnbaum, 1968). It involves estimations of either item and thresholds difficulties or discrimination. We utilized the maximum likelihood estimator with robust standard errors (Rhemtulla, Brosseau-Liard, & Savalei, 2012). The model comparison is based on the global fit statistics, such as Akaike information criterion (AIC) and Bayes information criterion (BIC, classic, and sample size adjusted versions). Pokropek, Marks, Borgonovi, Koc, and Greiff (2022) recommend relying on AIC as it is more appropriate for incomplete test design (especially when students perform different sets of items) and not susceptible to overfitting.

For studying developmental differentiation, such bifactor metric as explained common variance (*ECV*) was utilized. It demonstrates the proportion of general factor in the amount of total variance (general factor + specific factors). If its empirical value exceeds 0.70, a predominantly unidimensional structure is observed in the data (Rodriguez, Reise, & Haviland, 2016). Also, explained variance by a specific factor was implemented (ECV_{SS}). It specifies the proportion of a specific factor in the common variance accounted for by the general and specific factors. Finally, we utilized the Omega reliability coefficients (general and hierarchical). The general version describes the amount of reliable variance due to the general or specific factor. According to Reise, Bonifay, and Haviland (2013), if its value exceeds 0.80 then the total score can be considered unidimensional. The hierarchical one is served for quantification of the extent to which subscale scores are not confounded by the general factor. The Omega Hierarchical is the percentage of systematic variance in raw total scores that can be attributed to individual differences in the general or specific factors (Dueber, 2017; Reise et al., 2013). To compare the explained common variance across grades, we computed the standard errors based on Andersson and Luo (2022). It allows us to compute the standard errors in reliance on asymptotic theory and implement the delta method. The latter assumes that variance of the confirmatory factor analysis (CFA) parameters might be approximated by the sum of their derivatives. The calculation of the partial derivatives for each factor loading took place with the following formula:

$$\frac{\partial ECV}{\partial \lambda} = \frac{1(\lambda \in \lambda_G)2\lambda\left(\sum_{j=1}^{J}\lambda_{Gj}^2 + \sum_{s=1}^{S}\sum_{j=1}^{J}\lambda_{sj}^2\right) - 2\lambda\left(\sum_{j=1}^{J}\lambda_{Gj}^2\right)}{\left(\sum_{j=1}^{J}\lambda_{Gj}^2 + \sum_{s=1}^{S}\sum_{j=1}^{J}\lambda_{sj}^2\right)^2},$$

where λ is a factor loading, λ_{Gj} is a factor loading for the general factor, λ_{sj} is a factor loading for a specific factor. Then, the derivatives of the corresponding factors averaged to obtain standard errors. Also, to investigate the developmental differentiation hypothesis, we computed the correlations across six PISA domains and estimated the mean and median correlation across grades. For studying relations between assessment domains, the latent factor correlations were used. It is worth noting that latent correlations allow for measurement error to be modeled and indicate true associations. Therefore, they are usually higher than observed correlations (Fleiss & Shrout, 1977).

To investigate the ability differentiation hypothesis, we used mixture modeling. This approach allows us to estimate both continuous and discrete variables simultaneously (Reynolds et al., 2010). We compared models with only two or three latent classes because for practical and research purposes a higher number of classes is unnecessary. Bayes method served as an estimator for these models because maximum likelihood computations are heavy due to many dimensions of integration. Numerical integration is needed when categorical outcomes are modeled with continuous latent variables (Muthén, 2010). For comparing models, a posterior predictive *p*-value (PPP) and entropy were used. PPP is a model fit estimate which indicates discrepancy (distance) between data and model parameters (Muthén & Asparouhov, 2012). As B. Muthén and T. Asparouhov underscore, an excellent-fitting model is expected to have a PPP value around 0.5 (Ibid.). Entropy is an indicator of classification quality and is derived from posterior class probabilities (Shin, No, & Hong, 2019). The closer it is to 1, the better (Celeux & Soromenho, 1996).

The statistical modeling was performed using MPlus version 7 (Muthén & Muthén, 1998-2012).

3. Results

3.1. Developmental differentiation

Table 4 contains fit statistics of unidimensional, multidimensional, and bi-factor models with six PISA subjects across grades. It is worth noting that the bi-factor version fits the data better.

Table 5 depicts the bifactor metrics of the common explained variance and Omega reliabilities. As we can see, the general factor explains at least 60% of the variance in students' responses, and its rate tends to rise with grade increase. The reading literacy dimension accounts for 4% to 13% of the variation in its items; variance explained by mathematical literacy varies between 5% and 9%, and by scientific literacy between 6% and 7%. Controlling for the effect of general factor, financial literacy explains 4–5% of dispersion of the respective items, global competence

Table 4
Global fit statistics for models including reading, mathematical, scientific, financial literacies, global competence, and creative thinking.

Grade	Model	N of free parameters	-2LL	AIC	BIC	sBIC
5th	Unidimensional	220	150,523	150,963	152,397	151,698
	Multidimensional	235	150,178	150,648	152,180	151,433
	Bifactor	304	149,681	150,289	152,270	151,304
6th	Unidimensional	202	191,826	192,230	193,547	192,905
	Multidimensional	217	190,504	190,938	192,353	191,663
	Bifactor	295	189,143	189,733	191,655	190,718
7th	Unidimensional	230	209,791	210,251	211,750	211,019
	Multidimensional	244	208,752	209,240	210,831	210,055
	Bifactor	321	207,758	208,400	210,492	209,472
8th	Unidimensional	217	178,784	179,218	180,632	179,943
	Multidimensional	232	178,006	178,471	179,982	179,245
	Bifactor	302	177,722	178,326	180,294	179,334

Note. -2LL – -2 x log-likelihood, AIC – Akaike information criterion, BIC – Bayesian information criterion, sBIC – sample size adjusted Bayesian information criterion.

Table 5
Bi-factor metrics for models including reading, mathematical, scientific, financial literacies, global competence, and creative thinking.

Grade	Variable	ECV	Omega	Omega hierarchical & subscale
	g	0.63	0.91	0.86
	RL	0.08	0.67	0.26
	ML	0.08	0.67	0.09
5th	SL	0.07	0.68	0.20
	CT	0.05	0.57	0.29
	GC	0.04	0.55	0.11
	FL	0.04	0.67	0.09
	g	0.60	0.92	0.84
	RL	0.13	0.83	0.24
	ML	0.07	0.76	0.13
6th	SL	0.07	0.71	0.22
	CT	0.06	0.54	0.40
	GC	0.02	0.61	0.18
	FL	0.05	0.77	0.13
	g	0.64	0.92	0.86
	RL	0.10	0.85	0.18
	ML	0.09	0.74	0.22
7th	SL	0.06	0.69	0.24
	CT	0.04	0.43	0.24
	GC	0.02	0.60	0.14
	FL	0.05	0.75	0.18
	g	0.70	0.92	0.87
	RL	0.04	0.70	0.08
	ML	0.05	0.61	0.19
8th	SL	0.06	0.62	0.13
	CT	0.08	0.67	0.37
	GC	0.03	0.72	0.05
	FL	0.04	0.68	0.09

describes 2–4% of the variation, and 4% to 8% of the variance are attributed to domain-specific creative thinking-factor. According to Omega values, domain-general dimension is measured more reliably than the domain-specific factors. The omega hierarchical for g-factor is higher than 0.80.

The standard error for ECV in 5th grade is 1.27%, in 6th grade it is 1.21%, in 7th grade it equals 1.15%, and in 8th grade it is 1.11%. Fig. 1 presents bar plots with 95% confidence interval error bars for ECV in each grade. The confidence intervals intersect in grades 5–7. At the same time, the confidence interval for ECV in grade 8 does not overlap with others.

Fig. 2 depicts the variance explained by the domain-specific factors and a linear approximation of its change (the confidence intervals omitted for clarity). Scientific literacy can be approximated by linear change across grades indicating a slight decline in effects but not significant. Financial literacy demonstrates a stagnation. Creative thinking and mathematics literacy in our study cannot be described by linear trends, the polynomial function of degree three fits them better. Reading literacy and global competence require the polynomial function of degree two.

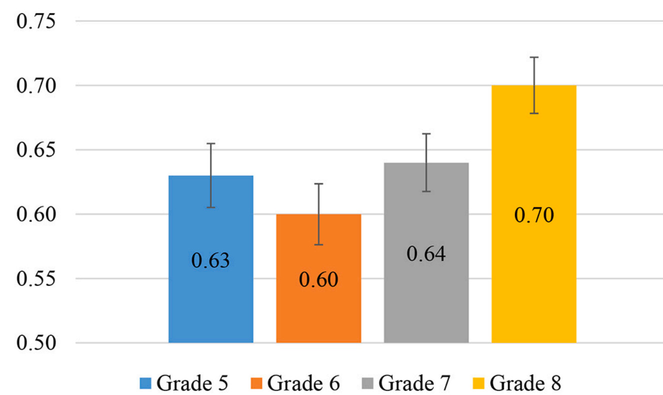

Fig. 1. Bar chart with 95% confidence interval error bars for ECV across grades.

Appendix A at the end of this article contains tables of bivariate correlations across six PISA subjects. In year 5, the strongest links are reading literacy with global competence and financial literacy, mathematics literacy with financial literacy, and global competence with scientific literacy and financial literacy (all correlations are higher than 0.80). The weakest one is reading literacy with scientific literacy (r = 0.39, $p < .01$). The mean correlation is 0.69 (Me = 0.69). In year 6, the strongest association is between financial literacy and global competence (r = 0.87, $p < .01$). The links of creative thinking with mathematics literacy and financial literacy are the weakest (r = 0.29, $p < .01$ and r = 0.36, $p < .01$ respectively). The mean correlation is 0.60 (Me = 0.65). In year 7, the strongest relations are reading literacy with global competence and of global competence with financial literacy (all correlations are higher than 0.80). The links between mathematics literacy and creative thinking and between financial literacy and creative thinking are the weakest (r = 0.47, $p < .01$ and r = 0.48, $p < .01$ respectively). The mean correlation is 0.65 (Me = 0.68). In year 8, the strongest correlations are: reading literacy with global competence, financial literacy with scientific literacy and global competence (all associations are equal or higher than 0.88). The links of creative thinking with mathematics literacy and global competence are the weakest (r = 0.45, $p < .01$ and r = 0.50, $p < .01$ respectively). The mean correlation is 0.72 (Me = 0.74).

Fig. 3 contains graphical representation of the mean and median correlations across grades and their respective linear approximations. The mean correlations vary between 0.60 and 0.72, and the median ones between 0.62 and 0.74. The patterns do not approximate with linear trends and follow the polynomial function of degree two. The associations in grades 6 and 7 are significantly smaller than in grade 8 and do not differ from correlations in grade 5. The linear increasing trend is observed only from 6th to 8th grades.

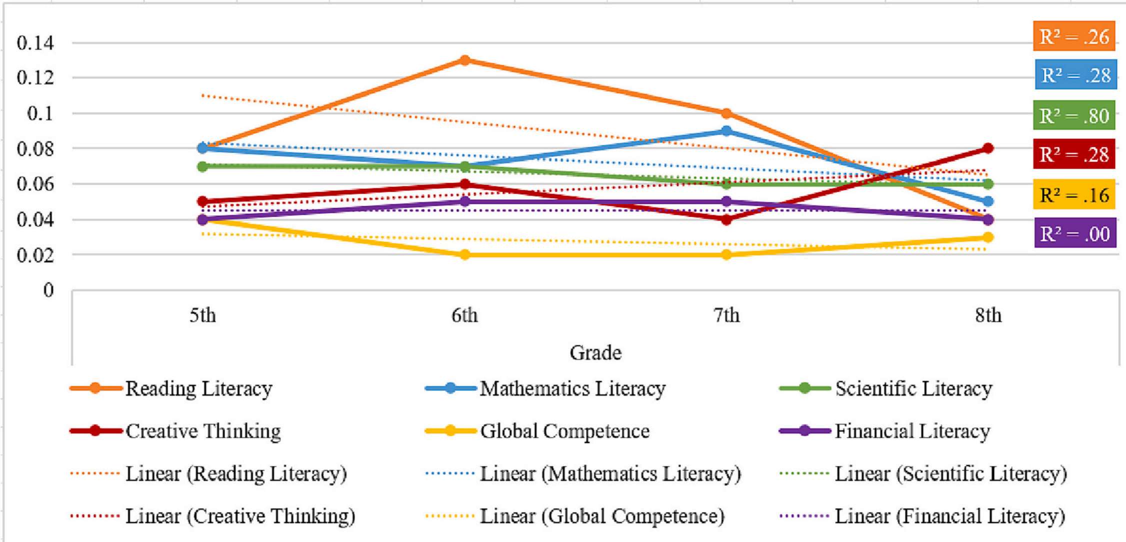

Fig. 2. Explained variance by a specific factor across grades and linear approximations (six PISA domains).

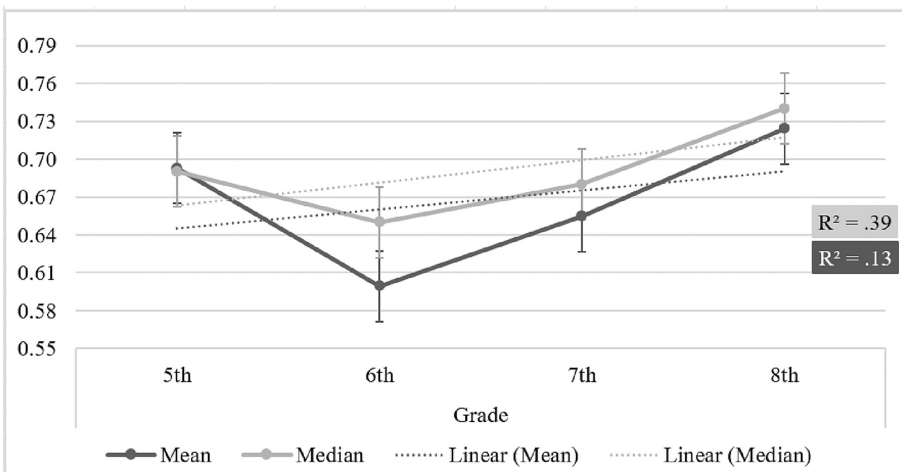

Fig. 3. Mean and median correlations (with 95% confidence intervals) of the PISA domains across grades and their linear approximations.

3.2. *Ability differentiation*

Table 6 presents fit statistics for mixture models. The posterior predictive *p*-value is close to 0.50, entropy exceeds 0.90 in all grades. It is worth noting that there are no substantive differences between two-class and three-class solutions. Consequently, for the sake of clarity, we end up with 2 latent classes and the following description will be of the two-class decision.

Table 7 contains the mean values and variances of g-factor across two latent classes obtained on the previous step. The averages for the first latent class are centered at 0. All factor values in latent class 2 are significantly higher than zero. Moreover, variances of g-factor within grades are equal.

Table 6
Fit statistics of the mixture models.

Grade	Model	Posterior predictive p-value (PPP)	Entropy
5th	2 classes	0.46	0.91
	3 classes	0.47	0.94
6th	2 classes	0.39	0.91
	3 classes	0.41	0.94
7th	2 classes	0.37	0.91
	3 classes	0.42	0.95
8th	2 classes	0.44	0.94
	3 classes	0.44	0.92

Table 7
Unstandardized mean scores and variances of g-factor across latent classes.

Grade	M (Var)		Cohen's d
	Class 1	Class 2	
5th	0.00 (0.06)	0.11** (0.06)	0.45
6th	0.00 (0.07)	0.07* (0.07)	0.26
7th	0.00 (0.06)	0.21** (0.06)	0.86
8th	0.00 (0.05)	0.26** (0.05)	1.16

Note. ** $p < .01$, * $p < .05$.

Fig. 4 depicts the proportion of variance explained by g-factor in low versus high-ability students across grades. The latent class having lower estimate of g-factor is provisionally indicated as low and the second class, as high. In grade 5, the rate of variance explained by g-factor in the low-level group is not significantly different from the high-level one. In contrast, in grades 6–8, the dominance of g-factor in lower-level students is more pronounced than in higher-level latent classes. The

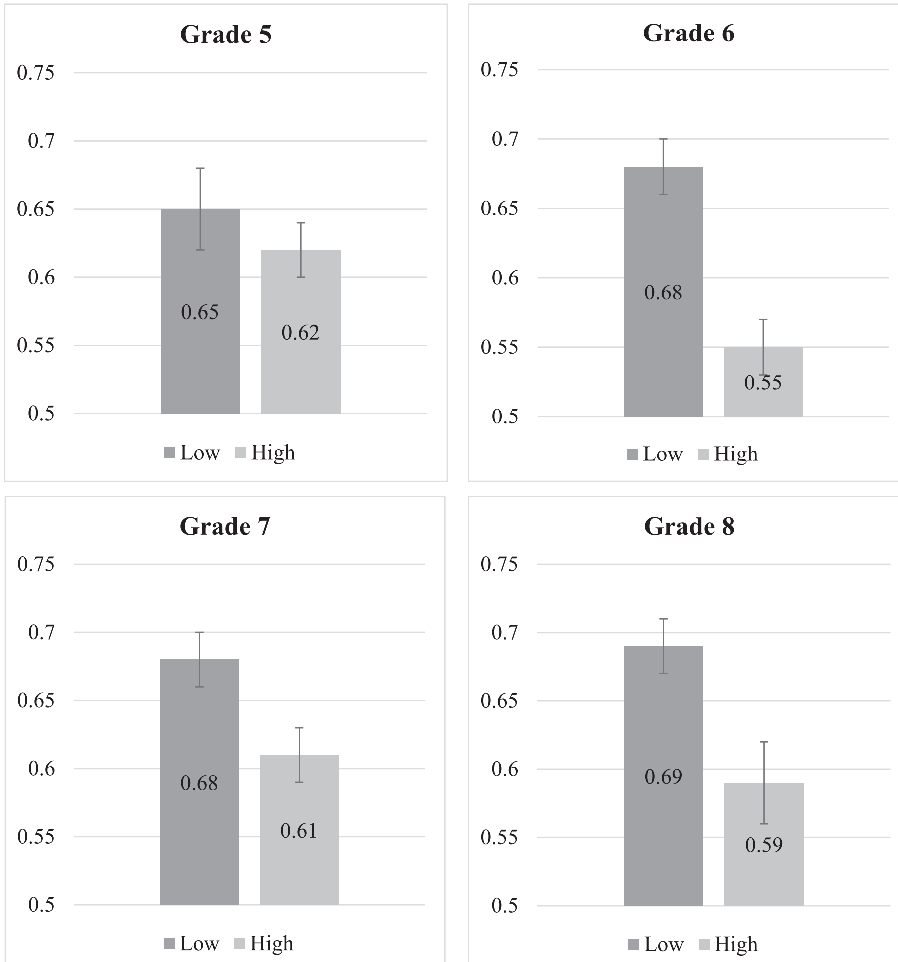

Fig. 4. Variance explained by g-factor in low- versus high-ability latent groups of students (with 95% confidence intervals).

difference in explained variance varies between 7% (grade 7) and 13% (grade 6).

Fig. 5 shows the proportion of variance explained by specific factors in low versus high-ability students across grades. In grade 5, the differences between lower-ability and higher-ability groups are significant in reading literacy and global competence, although the difference in the global competence results would not be significant at a confidence level of 99%. In grade 6, the differences between groups are significant in reading, mathematics, science, and financial literacies. In grade 7, the only significant difference is in reading literacy. In contrast, in grade 8, the significant differences are in mathematics and science.

4. Discussion

This study is dedicated to examining the factor structure of the PISA-like tests and the effect of the general intelligence across four school grades – from 5th to 8th. We revealed that the data is better described with the bi-factor model that accounts for a domain-general factor and six specific ones. We confirmed partially all stated hypotheses. Regarding H_1 (that g-factor will explain most of the variance in students' responses), we indicated that the variance explained varied between 60% and 70% depending on the grade. As for H_2 (that the dominance of the general factor will increase from 5th to 8th grades), we demonstrated that from 5th to 8th grades, the proportion of g-factor in the variance of the test results increased from 63% to 70%. However, the differences in the rate of g-factor across 5 to 7 grades are not significant. Only an increase from grade 7 to grade 8 is noticeable. Concerning H_3 (that in lower-ability students g-factor will explain more variance compared to higher-ability students in each sampled grade), we found that in grades 6–8 the difference in the proportion of g-factor between lower- and higher-ability groups of students is significant. In contrast, in grade 5, the difference is not statistically significant.

Although regarding developmental differentiation, our findings are inconsistent, they correspond to Pokropek, Marks, and Borgonovi (2022); Pokropek, Marks, Borgonovi, Koc, and Greiff (2022) results. In their study, the general factor explained >70%–80% of the variance in the students' responses. In addition, Breit et al. (2020) in their two studies supported the idea of no age differentiation during adolescence. Also, our findings coincide with results of Breit et al.'s (2022) systematic review by showing no age effect from 5 to 7 grade and age dedifferentiation from 7 to 8 grade. Our calculations indicate 60% to 70% of the variance explained by g-factor. Moreover, under the same test design used in different grades, we found a 7% increase in the contribution of general intelligence from grade 5 to grade 8. This result might be explained by improvements in working memory which is strongly associated with intelligence. For example, Huizinga, Dolan, and Van der Molen (2006) found that their participants did not reach the adult level of performance in working memory tasks until the age of 15. On a representative sample of U.S. households, Ahmed, Ellis, Ward, Chaku, and Davis-Kean (2022) revealed that 14–15-year-olds overcame 11–13-year-olds in the backward digit span test (which is a measure of working memory). At the same time, in our country, eighth-graders are usually 14 or 15 years old, and during the testing procedure they could demonstrate an adult level of working memory span that helped them be successful in a broader range of items compared to students from grades 5–7. The improvement in working memory might result from structural

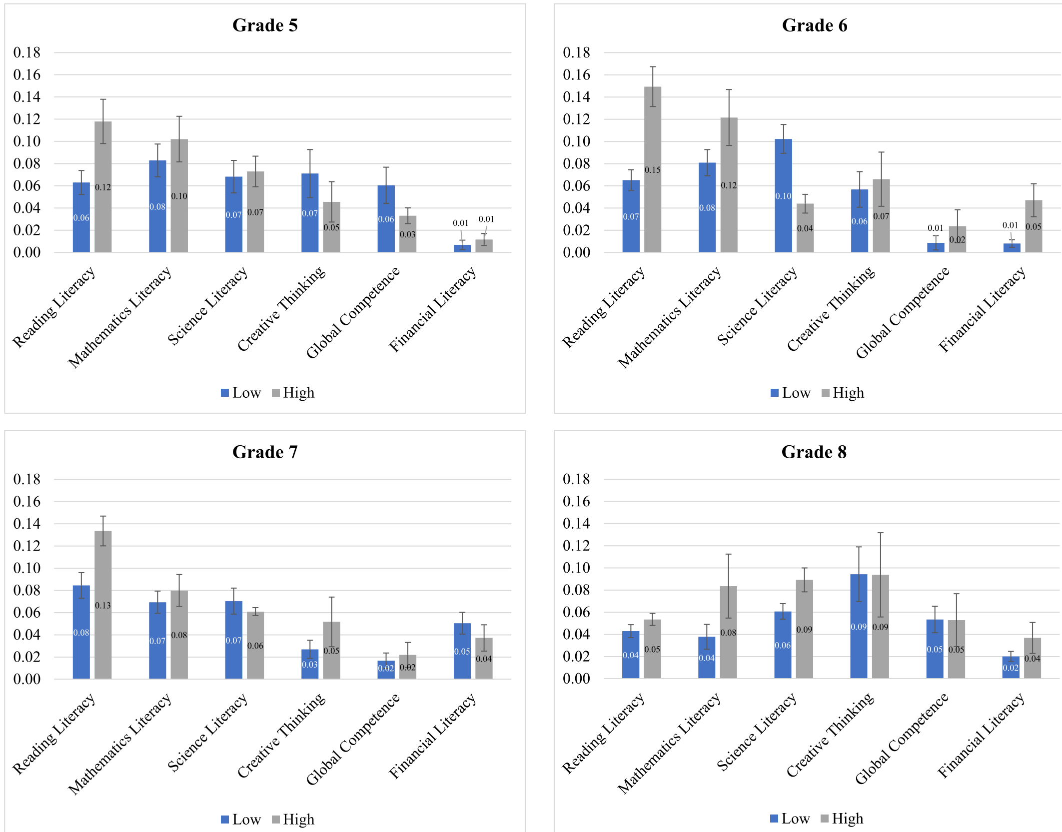

Fig. 5. Variance explained by specific factors in low- versus high-ability latent groups of students (with 95% confidence intervals).

reorganization of its mechanisms in the brain to adult mode that relies more on anterior sites (Gómez et al., 2018).

Speaking of dimensionality of the data structure, we could note that according to Omega coefficients (both general and hierarchical) the data demonstrate unidimensionality. Furthermore, if we calculate differences between Omega general and Omega hierarchical, we will find the proportion of the variance attributed to multidimensionality (Rodriguez et al., 2016). In grades 5 and 8 the difference is 5%, in grade 6 it is 8%, and in grade 7 it equals 6%. It means that the raw total scores obtained from the assessments might be interpreted as reliable reflections of the general factor. If we look at the common explained variance (ECV), the picture is not so clear. Although the general factor in each grade accounts for over half the common variance, the data in grades 5 to 7 are not considered purely unidimensional because their ECV is significantly smaller than the cut-off value of 0.70. However, bearing in mind the low estimates of Omega subscale for specific factors, it might be reasonable to extract only one raw score to investigate individual differences in intelligence. In grade 8, in terms of ECV, the results of the testing procedure are unidimensional.

We also found that the mean and median correlations of test components do not fit well with the linear function if we account for every grade in our study; they shape the U-pattern. Nevertheless, the positive trend is present from grade 6 to grade 8. This evidence is in partial accordance with age dedifferentiation. On a representative sample of German elementary school children (5- to 12-year-olds), Breit, Brunner, and Preckel (2021) declared no evidence of systematic age dedifferentiation. They underscore that age differentiation might occur in younger primary school students and decline by the age of 12. In our case, the difference in the proportion of variance attributable to the general intelligence factor across grades 5, 6, and 7 is insignificant (no differentiation or dedifferentiation).

Also, we revealed that only the contribution of scientific literacy into the variance of students' responses from grade 5 to grade 8 might be approximated with the linear trend indicating that its proportion slightly diminishes. The rate of financial literacy does not change; it stagnates at about 4%. Global competence seems not to change; however, it fits the linear function badly (it fluctuates around 2–4% of the specific variance) as well as mathematics literacy, which is described better with a polynomial function of degree three. Reading literacy demonstrates highest contribution in test results in grades 5–7. As opposed to other test components, creative thinking exhibits a minor increase from 7th to 8th grade. It may be reasonable since creativity develops non-linearly, especially in adolescence (Barbot & Tinio, 2015). Also, it might be in accordance with the idea of Kleibeuker, De Dreu, and Crone (2013) who showed an increase in visuospatial divergent and verbal thinking from 12/13 to 15/16 years. According to our analysis based on the multidimensional model, creative thinking usually demonstrates lower correlations with other domains across grades. On the one hand, this may be explained by the lower reliability of creative thinking measures, because of the peculiarities of this area of assessment, a student cannot perform many tasks in a given time. Therefore, this affected the measurement precision. On the other hand, this might be evidence of a slightly higher specificity of creativity compared to other domains. For instance, Baer (2016) in many studies pointed out that creative thinking is either content or task specific. Moreover, Kaufman and Baer (2004) revealed that self-perceived creativity is associated with other domains (such as writing, art, even science), except mathematics. In our study, the lowest correlations are exactly between mathematical literacy and creative thinking from 6th to 8th grade. Contrary to their expectations, de Vink, Willemsen, Lazonder, and Kroesbergen (2022) on a sample of primary school children found no link between divergent thinking and mathematics performance announcing greater role of convergent thinking. In addition, our mathematical literacy items contain a lot of algorithmicity which is contrary to the nature of creativity. At the same time, mathematical creativity, which is actively being studied, requires a higher expertise in mathematics (Grégoire, 2016) that middle school students can barely demonstrate.

As for ability differentiation, we found that the proportion of g-factor is a function of ability level in grades 6 to 8. In grade 5, the ability differentiation hypothesis is not confirmed. These findings intersect Breit et al. (2020). The scholars found that ability differentiation significantly depended on age: younger adolescents (10–12-year-olds) demonstrated a more definite effect of ability dedifferentiation; at the same time, older adolescents experienced ability differentiation. In our study, 12–14-year-old students are influenced by ability differentiation, while 11-year-olds (grade 5) are not. Also, they showed that in older adolescents (about 17 years old) the ability differentiation is greater than in younger adolescents. We could not support the idea of a moderating role of age in ability differentiation since in our case the largest difference in the proportion of g-factor between low and high ability groups was in grade 6 (diff. = 13%), while in grades 7 and 8 the differences were smaller (7% and 10% respectively).

In addition, we revealed a specific ability differentiation in grades 5 to 7, especially in reading literacy. Students from higher ability latent classes have a more pronounced proportion of variance explained by reading literacy. It might indicate that high ability students tend to develop specific reading skills that help them be more successful in PISA-like tests; potentially, not only in reading literacy tests, because other assessment domains may require higher reading competencies (OECD, 2019c). Moreover, this finding corresponds to results of Breit et al. (2021) who encountered differentiation in verbal reasoning on a sample of students from grades 1 to 4. It is worth noting that phonological processing skills are associated and predictive of reading abilities and reading comprehension (Cain & Oakhill, 2006; McBride-Chang & Manis, 1996). Our results might indicate that the trend of verbal reasoning differentiation continues in secondary school (at least when performing tasks on reading literacy).

4.1. Implications

Some scholars underline that ability differentiation might be observed because intelligence tests are usually designed for low and average ability examinees (Breit, Brunner, & Preckel, 2019). It means that in the presence of a higher proportion of easy items respondents with higher levels of intelligence cannot demonstrate their full potential and complexity of their thought; so, it leads to a g-saturation in lower ability people. In our case, on average, item difficulty is higher than students' ability on the logit scale. Appendix B includes Wright maps of students' ability and item difficulties from the calibration procedure that helps visually compare two distributions. Students' results are placed on the left and the item difficulty parameters, on the right. In these maps, the mean item difficulty is centered at 0, and students' ability is freely estimated. It is also evident from Table 1 which contains ability distribution statistics that substantiates a skewness of the students' results toward lower values. Ability distributions in all grades are mostly below 0. We found ability differentiation in grades 6 to 8, so, potentially, this effect could be explained less by the difficulty of the test items and more by other factors (such as motivation or interest).

Also, as posited by mutualism theory, in human development periods of integration and differentiation may coexist (Van Der Maas et al., 2006). If we rely on this postulate and combine it with our results, we can assume that in younger adolescents (grades 5–7) there is no developmental differentiation effect while ability differentiation mostly persists (except in students of grade 5). Furthermore, younger adolescents may experience differentiation in specific abilities, especially in reading literacy and comprehension. In contrast, older adolescents might be characterized by general ability differentiation and developmental dedifferentiation (integration); specific ability differentiation is not highly pronounced in them because the rate of g-factor rises. It seems, the dynamic mutualism theory explains the findings regarding ability and developmental differentiation better than other theoretical models.

In addition, because there is little research on intelligence and specific abilities based on PISA and PISA-like tests, explanatory approaches of the results obtained in such studies have not yet been elaborated. At the same time, such research is extremely important, especially for educators. The use of contextual (school-oriented) tests to diagnose general and specific abilities in order to better understand their formation can help all agents of the educational process to adjust to it, achieving better results. We do not pretend to replace the school psychologist and their cognitive testing duties with PISA-like tests. Nevertheless, in our point of view, the combination of psychological and pedagogical assessments can be beneficial and bring new insights.

4.2. Limitations

The present study is not free of limitations. First, our study was conducted on students in grades 5–8 from different regions. Even though each of the samples is representative, we cannot transfer the results to all students in the Russian Federation because the samples represent only pupils from their respective regions. Also, different regions could give different results manifested in fluctuations in correlations and explained variance. Participants from different grades but in the same regions might be needed in future research. Second, the results were obtained under cross-sectional design, and we observed the changes in the contribution of the general factor on different samples. This approach might not reflect the real advance of students' abilities. Longitudinal studies are necessary for establishing a more ecologically valid developmental transition of the g-factor effect. Third, our test design is distinct from that used in PISA, both in content and procedure. Consequently, we cannot unequivocally compare our findings with those of studies conducted on PISA data. Nevertheless, the tendencies pointed out in our study and observed in others are similar. Fourth, in our test variants, there were a disproportionate number of items measuring different literacies (especially in mathematics and reading). Potentially, this could affect the rate of the general and specific factors. Fifth, because only 20 min were allotted to each of the blocks in the test variants, the six components were not measured highly reliably, which could bias the estimates of the proportion of the common and specific factors. Future studies may require other test designs (perhaps with a longer test procedure) to assess the contribution of common and specific factors more reliably. Finally, because different test variants were used in each grade and there was no vertical scaling, the comparisons across grades are approximate. Even though the tests were designed in reliance on the same conceptual blueprint, the assessment results are best described with the same psychometric model (bifactor), generally speaking, the students' results across grades are not presented on the common scale. Moreover, because of a lack of vertical scaling we cannot be completely sure that age dedifferentiation is necessarily due to age-related cognitive changes and not to different test forms. Thereafter, vertical linking of test results is necessary to make conclusions about ability and developmental differentiation more credible.

5. Conclusion

Summarizing, the present study provided inconsistent evidence for an increasing contribution of the general intelligence factor from grades 5 to 8. The findings intersect with previous research on developmental dedifferentiation indicating a loss of cognitive specificity with age and maturation. The correlational analysis gave a contradictory picture, suggesting a level of specificity in grade 5 comparable to grade 8. Also, the study revealed ability differentiation of both general and specific factors. Younger adolescents with higher levels of intelligence have a larger proportion of the reading literacy component compared to their lower-level counterparts. Older adolescents experience mostly general ability differentiation. The results might be of practical importance. Students' cognitive abilities gain more generality as they grow older, and it could be said that the potential for teachers to cooperate in developing students' cognitive abilities can be exploited. For instance, integrated lessons where the same life problem is addressed from the perspectives of different scientific disciplines (e.g., mathematics, chemistry, and physics) may be offered. Also, it is possible to carry out projects, research activities, so that students could analyze and evaluate some issues from different scientific points of view. As for ability differentiation, the need for well-known differentiated instruction may follow from it. However, it is not only about differentiation in content, but also in the development of skills. For instance, if lower-ability students rely more on general cognitive ability when solving problems and use little specific skills, it probably makes sense for them to train specific competencies.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

We are grateful to all our colleagues who developed and corrected items for the PISA-like tests.

Appendix A

Bivariate correlations with 95% confidence intervals of six PISA subjects (grade 5).

	RL	ML	SL	CT	GC
ML	0.73 [0.64; 0.82]				
SL	0.39 [0.17; 0.60]	0.60 [0.45; 0.74]			
CT	0.65 [0.54; 0.77]	0.64 [0.34; 0.94]	0.69 [0.59; 0.79]		
GC	0.82 [0.69; 0.94]	0.65 [0.51; 0.79]	0.62 [0.52; 0.71]	0.72 [0.63; 0.82]	
FL	0.84 [0.77; 0.92]	0.84 [0.76; 0.92]	0.74 [0.67; 0.81]	0.63 [0.56; 0.69]	0.83 [0.76; 0.91]

Note. RL – reading literacy, ML – mathematics literacy, SL – scientific literacy, FL – financial literacy, GC – global competence, CT – creative thinking.

Bivariate correlations with 95% confidence intervals of six PISA subjects (grade 6).

	RL	ML	SL	CT	GC
ML	0.56 [0.50; 0.63]				
SL	0.63 [0.56; 0.69]	0.68 [0.63; 0.74]			
CT	0.43 [0.36; 0.50]	0.29 [0.21; 0.36]	0.51 [0.42; 0.59]		
GC	0.73 [0.65; 0.81]	0.70 [0.63; 0.77]	0.65 [0.58; 0.71]	0.44 [0.34; 0.54]	
FL	0.72 [0.65; 0.79]	0.77 [0.71; 0.83]	0.65 [0.57; 0.72]	0.36 [0.27; 0.44]	0.87 [0.80; 0.93]

Note. RL – reading literacy, ML – mathematics literacy, SL – scientific literacy, FL – financial literacy, GC – global competence, CT – creative thinking.

Bivariate correlations with 95% confidence intervals of six PISA subjects (grade 7).

	RL	ML	SL	CT	GC
ML	0.61 [0.55; 0.67]				
SL	0.68 [0.62; 0.73]	0.65 [0.59; 0.71]			
CT	0.53 [0.46; 0.61]	0.47 [0.39; 0.55]	0.59 [0.49; 0.68]		
GC	0.85 [0.78; 0.91]	0.68 [0.59; 0.77]	0.70 [0.62; 0.78]	0.52 [0.44; 0.61]	
FL	0.75 [0.71; 0.80]	0.73 [0.66; 0.79]	0.75 [0.68; 0.82]	0.48 [0.39; 0.57]	0.83 [0.77; 0.89]

Note. RL – reading literacy, ML – mathematics literacy, SL – scientific literacy, FL – financial literacy, GC – global competence, CT – creative thinking.

Bivariate correlations with 95% confidence intervals of six PISA subjects (grade 8).

	RL	ML	SL	CT	GC
ML	0.74 [0.69; 0.78]				
SL	0.79 [0.73; 0.84]	0.72 [0.66; 0.79]			
CT	0.66 [0.61; 0.71]	0.45 [0.38; 0.52]	0.58 [0.50; 0.66]		
GC	0.88 [0.82; 0.94]	0.80 [0.74; 0.85]	0.70 [0.58; 0.82]	0.50 [0.41; 0.59]	
FL	0.81 [0.76; 0.86]	0.83 [0.78; 0.88]	0.91 [0.82; 1.00]	0.61 [0.49; 0.73]	0.88 [0.83; 0.93]

Note. RL – reading literacy, ML – mathematics literacy, SL – scientific literacy, FL – financial literacy, GC – global competence, CT – creative thinking.

Appendix B

Wright map of students' ability and item difficulty (grade 5).

```
         Person - MAP - Item
             <more>|<rare>
    3               +
                    |
                    |
                    |  RL_block_1_item_08
                    |
                    |
    2               +  FL_block_1_item_04  RL_block_2_item_08
                    |
                    |T RL_block_2_item_07
                    |  SL_block_2_item_03  FL_block_2_item_08
                 .  |
                 .  |  RL_block_2_item_06
    1            .  +  SL_block_1_item_05  FL_block_1_item_03
                       FL_block_2_item_07  RL_block_1_item_02
                 .  |S SL_block_1_item_07  FL_block_1_item_07
                       RL_block_2_item_01
                 .  |  GC_block_1_item_01  SL_block_2_item_08
                       RL_block_1_item_04  RL_block_1_item_06
                 .  |  GC_block_1_item_02  GC_block_2_item_01
                       SL_block_1_item_08  ML_block_2_item_05
                       FL_block_1_item_05  RL_block_1_item_07
                       RL_block_2_item_04
                 .  |  GC_block_2_item_02  SL_block_1_item_02
                       SL_block_1_item_04  SL_block_2_item_04
                       SL_block_2_item_06  ML_block_1_item_03
                       RL_block_2_item_05  RL_block_2_item_09
                .# T|  SL_block_2_item_02  SL_block_2_item_09
                       ML_block_2_item_02  RL_block_1_item_09
                       RL_block_2_item_03
    0          .##  +M SL_block_1_item_01  SL_block_1_item_03
                       SL_block_2_item_01  CT_block_2_item_01  CT_block_2_item_03
                       FL_block_2_item_03  RL_block_1_item_01
                       RL_block_1_item_05
              .###  |  GC_block_2_item_03  RL_block_2_item_02
                       RL_block_2_item_10
             .####  |  SL_block_1_item_09  SL_block_2_item_05
                       CT_block_1_item_02  ML_block_1_item_02
                       RL_block_1_item_10
        .######### S|  GC_block_1_item_05  SL_block_1_item_06
                       SL_block_2_item_07  CT_block_1_item_04
                       CT_block_1_item_05  CT_block_2_item_02
                       ML_block_1_item_01  ML_block_1_item_04
                       ML_block_2_item_04  ML_block_2_item_06
                       FL_block_2_item_06  RL_block_1_item_03
          .#######  |  GC_block_1_item_03  ML_block_1_item_06
                       ML_block_2_item_01  ML_block_2_item_03
                       FL_block_1_item_02  FL_block_2_item_01
                       FL_block_2_item_02
       .##########  |S GC_block_2_item_04  CT_block_1_item_03
                       CT_block_2_item_04  FL_block_2_item_04
   -1  .##########  +  ML_block_1_item_05
     #############  |  FL_block_1_item_08
     .############ M|  GC_block_2_item_05  CT_block_1_item_01
       .##########  |
        .#########  |T GC_block_1_item_04  FL_block_1_item_06
         .########  |  FL_block_2_item_05
   -2      .######  +
            .##### S|  FL_block_1_item_01
           .######  |
                .#  |
               .##  |
              .###  |
   -3              . T+
                 .  |
                 .  |
                 .  |
                .#  |
                 .  |
   -4               +
                 .  |
                 .  |
                    |
                    |
                    |
   -5            .  +
             <less>|<frequ>
  EACH "#" IS 197. EACH "." IS 1 TO 196
```

Wright map of students' ability and item difficulty (grade 6).

```
           Person - MAP - Item
              <more>|<rare>
    2                +
                     |T
                     |
                     |  SL_block_1_item_01  SL_block_1_item_03
                        SL_block_1_item_08
                     |  SL_block_1_item_02  SL_block_1_item_04
                        ML_block_2_item_08
                     |  SL_block_1_item_07  CT_block_1_item_02
                        ML_block_1_item_08
    1              . +S GC_block_2_item_02  CT_block_1_item_04
                        ML_block_1_item_05  FL_block_1_item_04
                   . |  GC_block_1_item_01  GC_block_2_item_03
                        SL_block_1_item_05  SL_block_1_item_06
                        SL_block_2_item_05  SL_block_2_item_09
                        ML_block_1_item_03  ML_block_1_item_06
                        ML_block_1_item_07
                   . |  SL_block_1_item_10  ML_block_2_item_01
                        ML_block_2_item_06  FL_block_1_item_01
                        RL_block_1_item_10  RL_block_2_item_01
                        RL_block_2_item_10
                   . |  SL_block_2_item_02  ML_block_2_item_03
                        ML_block_2_item_07  RL_block_1_item_02
                        RL_block_2_item_11
                   . T| GC_block_1_item_03  GC_block_1_item_05
                        SL_block_2_item_03  SL_block_2_item_07
                        ML_block_1_item_01  ML_block_2_item_02
                        FL_block_2_item_02  FL_block_2_item_08
                        RL_block_1_item_04
                  .# |  GC_block_1_item_02  GC_block_2_item_05
                        ML_block_2_item_05  RL_block_1_item_07
                        RL_block_1_item_11  RL_block_2_item_07
    0           .### +M GC_block_1_item_04  SL_block_2_item_01
                        CT_block_2_item_04  ML_block_2_item_04
                        FL_block_1_item_03  RL_block_1_item_01
              .##### |  SL_block_2_item_06  SL_block_2_item_10
                        CT_block_2_item_03A CT_block_2_item_03B
                        RL_block_2_item_04
            .####### |  SL_block_2_item_04  SL_block_2_item_08
                        CT_block_1_item_03  FL_block_1_item_08
                        RL_block_1_item_09
          .######### S| SL_block_1_item_09  CT_block_2_item_02
                        FL_block_2_item_05  RL_block_1_item_03
         .########## |  GC_block_2_item_01  ML_block_1_item_02
                        ML_block_1_item_04  FL_block_1_item_02
                        RL_block_2_item_02  RL_block_2_item_03
                        RL_block_2_item_05
        .########### |  GC_block_2_item_04  CT_block_2_item_01
                        FL_block_2_item_01  FL_block_2_item_03
                        RL_block_1_item_05
   -1  .############ +S FL_block_2_item_06  RL_block_2_item_09
          .######### M| CT_block_1_item_01
        .########### |  FL_block_2_item_07
       .############ |  RL_block_2_item_06
             .###### |  RL_block_1_item_06  RL_block_2_item_08
            .####### |T FL_block_1_item_05  FL_block_1_item_06
   -2        .###### S+ RL_block_1_item_08
               .#### |
                 .## |  FL_block_2_item_04
                 .## |
                  .# T|
                  .# |
   -3              . +
                   . |
                   . |
                     |
                   . |
                   . |  FL_block_1_item_07
   -4              . +
                   . |
                     |
                   . |
                     |
                   . |
   -5              . +
                     |
                     |
                     |
                     |
                     |
   -6              . +
              <less>|<frequ>
EACH "#" IS 163. EACH "." IS 1 TO 162
```

Wright map of students' ability and item difficulty (grade 7).

```
         Person - MAP - Item
             <more>|<rare>
    2                +T
                     |
                     |  SL_block_1_item_07
                     |  SL_block_1_item_08  CT_block_1_item_04
                        ML_block_1_item_06  ML_block_2_item_05
                     |  SL_block_1_item_01  ML_block_1_item_07
                        ML_block_1_item_08
                     |  SL_block_1_item_06  SL_block_1_item_03
                        SL_block_2_item_10  CT_block_1_item_02
    1                +S GC_block_1_item_02  SL_block_1_item_09
                        SL_block_2_item_04  ML_block_1_item_03
                  .  |  SL_block_1_item_02  ML_block_2_item_04
                        FL_block_1_item_04  RL_block_2_item_10
                  .  |  GC_block_2_item_01  GC_block_1_item_03
                        GC_block_1_item_05  SL_block_1_item_05
                        ML_block_1_item_05  ML_block_2_item_01
                        ML_block_2_item_03  RL_block_1_item_10
                  .  |  SL_block_1_item_10  SL_block_2_item_02
                        ML_block_2_item_02  ML_block_2_item_06
                        ML_block_2_item_07  FL_block_1_item_01
                        RL_block_2_item_01
                  . T|  GC_block_2_item_05  SL_block_2_item_07
                        SL_block_2_item_08  ML_block_2_item_08
                        FL_block_2_item_08  RL_block_1_item_02
                        RL_block_1_item_07  RL_block_2_item_07
                 .#  |  GC_block_2_item_02  GC_block_2_item_03
                        ML_block_1_item_01  FL_block_2_item_02
                        RL_block_1_item_04
    0           .##  +M SL_block_2_item_05  CT_block_2_item_04
                        FL_block_1_item_03  RL_block_1_item_01
                        RL_block_2_item_04  RL_block_2_item_11
              .####  |  GC_block_2_item_04  SL_block_2_item_06
                        SL_block_2_item_09  CT_block_1_item_03
                        FL_block_1_item_08  RL_block_2_item_05
            .######  |  SL_block_2_item_01  ML_block_1_item_04
                        RL_block_1_item_11
         .######### S|  SL_block_1_item_04  SL_block_2_item_03
                        CT_block_2_item_03  FL_block_2_item_05
                        RL_block_1_item_03  RL_block_1_item_05
                        RL_block_2_item_02
        .##########  |  GC_block_1_item_01  CT_block_2_item_01
                        CT_block_2_item_02  FL_block_2_item_01
                        FL_block_1_item_02  RL_block_1_item_09
        .##########  |  FL_block_2_item_03  RL_block_2_item_03
   -1 .############  +S GC_block_1_item_04  CT_block_1_item_01
      .############  |  ML_block_1_item_02  RL_block_2_item_09
        .########## M|  FL_block_2_item_06
         .#########  |  FL_block_2_item_07  RL_block_2_item_06
          .########  |  RL_block_2_item_08
           .#######  |  FL_block_1_item_05  RL_block_1_item_06
   -2         .#### S+T FL_block_1_item_06
              #####  |
                .##  |  RL_block_1_item_08
               .###  |
               .###  |  FL_block_2_item_04
                  . T|
   -3             .  +
                 .#  |
                  .  |
                  .  |
                  .  |  FL_block_1_item_07
                  .  |
   -4             .  +
                  .  |
                     |
                  .  |
                     |
                  .  |
   -5             .  +
                     |
                  .  |
                     |
                     |
                     |
   -6             .  +
             <less>|<frequ>
 EACH "#" IS 84. EACH "." IS 1 TO 83
```

Wright map of students' ability and item difficulty (grade 8).

```
          Person - MAP - Item
               <more>|<rare>
   4              .  +
                     |
                     |
                     |
                     |
                     |
   3                 +
                     |
                     |
                     |  ML_block_2_item_05
                     |  RL_block_1_item_07
                     |  RL_block_1_item_01
   2                 +T SL_block_1_item_10  SL_block_2_item_03
                     |  FL_block_2_item_09
                     |  RL_block_1_item_02
                     |  SL_block_1_item_06
                  .  |  RL_block_1_item_04
                  .  |  SL_block_2_item_08  FL_block_2_item_04
   1              .  +S RL_block_1_item_05
                  .  |  SL_block_1_item_04  SL_block_1_item_05
                        SL_block_2_item_07  CT_block_1_item_02
                        ML_block_2_item_04  FL_block_1_item_07
                  .  |  SL_block_1_item_01  SL_block_1_item_08
                        CT_block_2_item_02  ML_block_2_item_02
                        RL_block_2_item_08
                  .  |  ML_block_2_item_03  ML_block_1_item_01
                        ML_block_1_item_02  ML_block_1_item_05
                        FL_block_1_item_04
                 .# T|  SL_block_1_item_07  CT_block_1_item_05
                        RL_block_2_item_07
                .##  |  GC_block_1_item_05  SL_block_1_item_09
                        SL_block_2_item_09  FL_block_2_item_08
                        RL_block_1_item_03
   0          .####  +M GC_block_1_item_01  GC_block_2_item_03
                        SL_block_2_item_01  CT_block_1_item_01
                        ML_block_1_item_03  RL_block_2_item_01
                        RL_block_2_item_04
            .######  |  RL_block_2_item_05
          .########  |  GC_block_1_item_02  SL_block_2_item_04
                        CT_block_1_item_04  CT_block_2_item_03
                        CT_block_2_item_04  ML_block_2_item_01
                        ML_block_1_item_04  FL_block_2_item_05
                        FL_block_1_item_02  RL_block_2_item_03
        .########## S|  GC_block_2_item_04  SL_block_1_item_02
                        SL_block_2_item_02  CT_block_2_item_01
                        FL_block_1_item_08  FL_block_1_item_10
                        FL_block_2_item_06  FL_block_2_item_10
                        RL_block_2_item_06
      .############  |  GC_block_1_item_04  GC_block_2_item_01
                        GC_block_2_item_02  SL_block_1_item_03
                        SL_block_2_item_05  CT_block_1_item_03
                        FL_block_2_item_03  RL_block_2_item_02
       .###########  |  FL_block_1_item_01  FL_block_1_item_05
                        FL_block_1_item_09
  -1  .############  +S GC_block_2_item_06  SL_block_2_item_06
                        FL_block_1_item_06  RL_block_1_item_06
      .############ M|  GC_block_1_item_03  GC_block_2_item_05
                        FL_block_1_item_03
      .############  |  GC_block_1_item_06
        .##########  |
         .#########  |
          .########  |  FL_block_2_item_07
  -2        .###### S+T
             .#####  |
             .#####  |
               .###  |  FL_block_2_item_01
                .##  |
                .## T|
  -3             .#  +
                  .  |
                  .  |
                  .  |
                  .  |
                  .  |
  -4              .  +  FL_block_2_item_02
                  .  |
                  .  |
                  .  |
                  .  |
                  .  |
  -5              .  +
                  .  |
                  .  |
                     |
                     |
                     |
  -6              .  +
               <less>|<frequ>
EACH "#" IS 248. EACH "." IS 1 TO 247
```

References

Ahmed, S. F., Ellis, A., Ward, K. P., Chaku, N., & Davis-Kean, P. E. (2022). Working memory development from early childhood to adolescence using two nationally representative samples. *Developmental Psychology, 58*(10), 1962–1973.

Andersson, B., & Luo, H. (2022). Impact of sampling variability when estimating the explained common variance. *Applied Psychological Measurement, 46*(4), 338–341.

Ardila, A. (2018). Is intelligence equivalent to executive functions? *Psicothema, 30*(2), 159–164.

Baer, J. (2016). Creativity doesn't develop in a vacuum. In B. Barbot (Ed.), *perspectives on creativity development. New Directions for Child and Adolescent Development, 151*, 9–20.

Barbot, B., & Tinio, P. P. (2015). Where is the "g" in creativity? A specialization–differentiation hypothesis. *Frontiers in Human Neuroscience, 8*, 1041.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612–637.

Baumert, J., Lüdtke, O., Trautwein, U., & Brunner, M. (2009). Large-scale student assessment studies measure the results of processes of knowledge acquisition: Evidence in support of the distinction between intelligence and student achievement. *Educational Research Review, 4*(3), 165–176.

Benedek, M., Beaty, R., Jauk, E., Koschutnig, K., Fink, A., Silvia, P. J., … Neubauer, A. C. (2014). Creating metaphors: The neural basis of figurative language production. *NeuroImage, 90*(100), 99–106.

Benedek, M., Jauk, E., Sommer, M., Arendasy, M., & Neubauer, A. C. (2014). Intelligence, creativity, and cognitive control: The common and differential involvement of executive functions in intelligence and creativity. *Intelligence, 46*, 73–83.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord, & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 396–479). Reading, MA: Addison-Wesley.

Blum, D., & Holling, H. (2017). Spearman's law of diminishing returns. A meta-analysis. *Intelligence, 65*, 60–66.

Boone, W. J., Staver, J. R., & Yale, M. S. (2013). *Rasch analysis in the human sciences*. Springer Science & Business Media.

Breit, M., Brunner, M., Molenaar, D., & Preckel, F. (2022). Differentiation hypotheses of intelligence: A systematic review of the empirical evidence and an agenda for future research. *Psychological Bulletin, 148*(7–8), 518–554.

Breit, M., Brunner, M., & Preckel, F. (2019). Spearman's law of diminishing returns and its implications for theories of general intelligence and intelligence testing. In D. J. McFarland (Ed.), *General and specific mental abilities* (pp. 340–358). Cambridge Scholars Publishing.

Breit, M., Brunner, M., & Preckel, F. (2020). General intelligence and specific cognitive abilities in adolescence: Tests of age differentiation, ability differentiation, and their interaction in two large samples. *Developmental Psychology, 56*(2), 364–384.

Breit, M., Brunner, M., & Preckel, F. (2021). Age and ability differentiation in children: A review and empirical investigation. *Developmental Psychology, 57*(3), 325–346.

Brinch, C. N., & Galloway, T. A. (2012). Schooling in adolescence raises IQ scores. *Proceedings of the National Academy of Sciences, 109*(2), 425–430.

Brunner, M. (2008). No g in education? *Learning and Individual Differences, 18*(2), 152–165.

Cain, K., & Oakhill, J. (2006). Profiles of children with specific reading comprehension difficulties. *The British Journal of Educational Psychology, 76*(4), 683–696.

Cattell, R. B. (1987). *Intelligence: Its structure, growth and action*. Elsevier.

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification, 13*(2), 195–212.

Cromley, J. G. (2009). Reading achievement and science proficiency: International comparisons from the programme on international student assessment. *Reading Psychology, 30*(2), 89–118.

D'Acunto, F., Hoang, D., Paloviita, M., & Weber, M. (2019). *IQ, expectations, and choice (No. w25496)*. National Bureau of Economic Research. Retrieved from https://www.nber.org/system/files/working_papers/w25496/w25496.pdf. Retrieved from.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135–168.

Dueber, D. M. (2017). Bifactor indices calculator: A Microsoft excel-based tool to calculate various indices relevant to bifactor CFA models. Retrieved from http://sites.education.uky.edu/apslab/resources/.

Evans, T. R., Hughes, D. J., & Steptoe-Warren, G. (2020). A conceptual replication of emotional intelligence as a second-stratum factor of intelligence. *Emotion, 20*(3), 507–512.

Feraco, T., & Cona, G. (2022). Differentiation of general and specific abilities in intelligence. A bifactor study of age and gender differentiation in 8-to 19-year-olds. *Intelligence, 94*, Article 101669.

Finke, M. S., Howe, J. S., & Huston, S. J. (2017). Old age and the decline in financial literacy. *Management Science, 63*(1), 213–230.

Fleiss, J. L., & Shrout, P. E. (1977). The effects of measurement errors on some multivariate procedures. *American Journal of Public Health, 67*(12), 1188–1191.

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science, 17*(2), 172–179.

Garrett, H. E. (1938). Differentiable mental traits. *The. Psychological Record, 2*, 259–298.

Gentner, D., & Maravilla, F. (2017). Analogical reasoning. In *The Routledge international handbook of thinking and reasoning* (pp. 186–203). Routledge.

Gómez, C. M., Barriga-Paulino, C. I., Rodríguez-Martínez, E. I., Rojas-Benjumea, M.Á., Arjona, A., & Gómez-González, J. (2018). The neurophysiology of working memory development: From childhood to adolescence and young adulthood. *Reviews in the Neurosciences, 29*(3), 261–282.

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

Gottfredson, L. S. (2005). Suppressing intelligence research: Hurting those we intend to help. In R. H. Wright, & N. A. Cummings (Eds.), *Destructive trends in mental health: The well-intentioned path to harm* (pp. 155–186). Routledge.

Grégoire, J. (2016). Understanding creativity in mathematics for improving mathematical education. *Journal of Cognitive Education and Psychology, 15*(1), 24–36.

Guilford, J. P. (1967). *The nature of human intelligence*. McGraw-Hill.

Guntersdorfer, I. R., & Golubeva, I. A. (2018). Emotional intelligence and intercultural competence: Theoretical questions and pedagogical possibilities. *Intercultural Communication Education, 1*(2), 54–63.

Gustavson, D. E., Reynolds, C. A., Corley, R. P., Wadsworth, S. J., Hewitt, J. K., & Friedman, N. P. (2022). Genetic associations between executive functions and intelligence: A combined twin and adoption study. *Journal of Experimental Psychology. General, 151*(8), 1745–1761.

Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory and Cognition, 15*(4), 332–340.

Huizinga, M., Dolan, C. V., & Van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia, 44*(11), 2017–2036.

Jauk, E., Benedek, M., Dunst, B., & Neubauer, A. C. (2013). The relationship between intelligence and creativity: New support for the threshold hypothesis by means of empirical breakpoint detection. *Intelligence, 41*(4), 212–221.

Kaufman, J. C., & Baer, J. (2004). Sure, I'm creative—But not in mathematics!: Self-reported creativity in diverse domains. *Empirical Studies of the Arts, 22*(2), 143–155.

Kaufman, J. C., Plucker, J. A., & Baer, J. (2008). *Essentials of creativity assessment*. John Wiley & Sons.

Kim, K. H. (2005). Can only intelligent people be creative? A meta-analysis. *Journal of Secondary Gifted Education, 16*(2–3), 57–66.

Kleibeuker, S. W., De Dreu, C. K., & Crone, E. A. (2013). The development of creative cognition across adolescence: Distinct trajectories for insight and divergent thinking. *Developmental Science, 16*(1), 2–12.

Kolmogorov, A. (1933). Sulla determinazione empirica di una lgge di distribuzione. *Giornale dell` Istituto Italiano degly Attuari, 4*(1), 83–91.

Lin, C. A., & Bates, T. C. (2022). Smart people know how the economy works: Cognitive ability, economic knowledge and financial literacy. *Intelligence, 93*, Article 101667.

Liu, Y., Yin, Y., & Wu, R. (2020). Measuring graduate students' global competence: Instrument development and an empirical study with a Chinese sample. *Studies in Educational Evaluation, 67*, Article 100915.

Lusardi, A., & Mitchell, O. S. (2014). The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature, 52*(1), 5–44.

Lusardi, A., Mitchell, O. S., & Curto, V. (2010). Financial literacy among the young. *Journal of Consumer Affairs, 44*(2), 358–380.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149–174.

McBride-Chang, C., & Manis, F. R. (1996). Structural invariance in the associations of naming speed, phonological awareness, and verbal reasoning in good and poor readers: A test of the double deficit hypothesis. *Reading and Writing, 8*(4), 323–339.

Miles, S., Howlett, C. A., Berryman, C., Nedeljkovic, M., Moseley, G. L., & Phillipou, A. (2021). Considerations for using the Wisconsin card sorting test to assess cognitive flexibility. *Behavior Research Methods, 53*(5), 2083–2091.

Mullis, I. V. S., & Martin, M. O. (2017). TIMSS 2019 assessment frameworks. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2019/frameworks/.

Muñoz-Murillo, M., Álvarez-Franco, P. B., & Restrepo-Tobón, D. A. (2020). The role of cognitive abilities on financial literacy: New experimental evidence. *Journal of Behavioral and Experimental Economics, 84*, Article 101482.

Muthén, B. (2010). Bayesian analysis in Mplus: A brief introduction. Retrieved from https://www.statmodel.com/download/IntroBayesVersion%203.pdf.

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods, 17*(3), 313–335.

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles: Muthén & Muthén.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., … Urbina, S. (1996). Intelligence: Knowns and unknowns. *The American Psychologist, 51*(2), 77–101.

OECD. (2012). *Results: Students and money financial literacy skills for the 21st century. Volume VI*. OECD Publishing. Retrieved from https://www.oecd.org/pisa/keyfindings/PISA-2012-results-volume-vi.pdf. Retrieved from.

OECD. (2019a). PISA 2021 Creative Thinking Framework (third draft). Retrieved from https://www.oecd.org/pisa/publications/PISA-2021-Creative-Thinking-Framework.pdf.

OECD. (2019b). *PISA 2018 assessment and analytical framework*. OECD Publishing. Retrieved from https://www.oecd-ilibrary.org/docserver/b25efab8-en.pdf?expires=1614758147&id=id&accname=guest&checksum=1091D0DAC117681FD3211B5329A0D1B0. Retrieved from.

OECD. (2019c). *PISA 2018 Results (Volume I). What students know and can do*. OECD Publishing. Retrieved from https://www.oecd-ilibrary.org/docserver/5f07c754-en.pdf?expires=1614758314&id=id&accname=guest&checksum=166FB427DAC867D0A92874F29AAF1071.

OECD. (2020). *PISA 2018 results (volume VI). Are students ready to thrive in an interconnected world?* OECD Publishing. Retrieved from https://www.oecd-ilibrary.org/docserver/d5f68679-en.pdf?expires=1614758237&id=id&accname=guest&checksum=D5BBAC650E8E46D9D808605186F326B7. Retrieved from.

Petrill, S. A., & Wilkerson, B. (2000). Intelligence and achievement: A behavioral genetic perspective. *Educational Psychology Review, 12*(2), 185–199.

Plomin, R., Fulker, D. W., Corley, R., & DeFries, J. C. (1997). Nature, nurture, and cognitive development from 1 to 16 years: A parent-offspring adoption study. *Psychological Science, 8*(6), 442–447.

Plomin, R., & Petrill, S. A. (1997). Genetics and intelligence: What's new? *Intelligence, 24*(1), 53–77.

Plomin, R., & Von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews. Genetics, 19*(3), 148–159.

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 Education & Psychology, 114*(5), 1121–1135.

Pokropek, A., Marks, G. N., Borgonovi, F., Koc, P., & Greiff, S. (2022). General or specific abilities? Evidence from 33 countries participating in the PISA assessments. *Intelligence, 92*, Article 101653.

Pokropek, A., & Sikora, J. (2015). Heritability, family, school and academic achievement in adolescence. *Social Science Research, 53*, 73–88.

R Core Team. (2020). R: A language and environment for statistical computing. Retrieved from https://www.R-project.org/.

Rasch, G. (1960). *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests*. Nielsen & Lydiche.

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment, 95*(2), 129–140.

Reynolds, M. R., Keith, T. Z., & Beretvas, S. N. (2010). Use of factor mixture modeling to capture Spearman's law of diminishing returns. *Intelligence, 38*(2), 231–241.

Rhemtulla, M., Brosseau-Liard, P.É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354–373.

Rietveld, C. A., Esko, T., Davies, G., Pers, T. H., Turley, P., Benyamin, B., & Koellinger, P. D. (2014). Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proceedings of the National Academy of Sciences, 111*(38), 13790–13794.

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*(5), 667–706.

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

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods, 21*(2), 137–150.

Saß, S., Kampa, N., & Köller, O. (2017). The interplay of g and mathematical abilities in large-scale assessments across grades. *Intelligence, 63*, 33–44.

Shelton, J. T., Elliott, E. M., Hill, B. D., Calamia, M. R., & Gouvier, W. D. (2009). A comparison of laboratory and clinical working memory tests and their prediction of fluid intelligence. *Intelligence, 37*(3), 283–293.

Shin, M., No, U., & Hong, S. (2019). Comparing the robustness of stepwise mixture modeling with continuous nonnormal distal outcomes. *Educational and Psychological Measurement, 79*(6), 1156–1183.

Sjøberg, S., & Jenkins, E. (2020). PISA: A political project and a research agenda. *Studies in Science Education*, 1–14.

Spearman, C. (1927). *The abilities of man*. Macmillan.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. Cambridge University Press.

Tucker-Drob, E. M. (2009). Differentiation of cognitive abilities across the life span. *Developmental Psychology, 45*(4), 1097–1118.

Van Der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review, 113*(4), 842–861.

de Vink, I. C., Willemsen, R. H., Lazonder, A. W., & Kroesbergen, E. H. (2022). Creativity in mathematics performance: The role of divergent and convergent thinking. *The British Journal of Educational Psychology, 92*(2), 484–501.

Wechsler, D. (2014). *Wechsler intelligence scale for children-fifth edition*. San Antonio, TX: NCS Pearson.

Whitley, E., Deary, I. J., Ritchie, S. J., Batty, G. D., Kumari, M., & Benzeval, M. (2016). Variations in cognitive abilities across the life course: Cross-sectional evidence from understanding society: The UK household longitudinal study. *Intelligence, 59*, 39–50.

Wickham, H., François, R., Henry, L., & Müller, K. (2022). dplyr: A grammar of data manipulation. https://dplyr.tidyverse.org. https://github.com/tidyverse/dplyr.

Willingham, D. T. (2021). *Why don't students like school? A cognitive scientist answers questions about how the mind works and what it means for the classroom*. John Wiley & Sons.

Woolley, G. (2010). Developing reading comprehension: Combining visual and verbal cognitive processes. *Australian Journal of Language and Literacy, 33*(2), 108–125.

Wright, B. D., Linacre, J. M., Gustafson, J. E., & Martin-Löf, P. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*(3), 370–371.