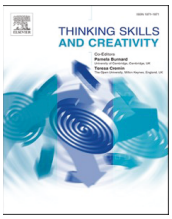
Thinking Skills and Creativity 59 (2026) 101963

Contents lists available at ScienceDirect

Thinking Skills and Creativity

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



Being female and being well-situated implies higher performance on creative thinking tests: Evidence across 62 countries from PISA 2022

B. Goecke [a,*], M. Benedek [b], J.G. Diedrich [c], B. Forthmann [d], S. Patzl [c,e], S. Weiss [f]

[a] Hector Research Institute of Education Sciences and Psychology, University of Tübingen, Tübingen, Germany
[b] University of Graz, Graz, Austria
[c] Center for International Student Assessment (ZIB), TUM School of Social Sciences and Technology, Technical University of Munich, Munich, Germany
[d] University of Münster, Münster, Germany
[e] University of Bamberg, Bamberg, Germany
[f] University of Hildesheim, Hildesheim, Germany

ARTICLE INFO

Keywords:
Creative thinking
PISA
Socio-economic status
Gender
Mathematics
Reading

ABSTRACT

Even in the time of artificial intelligence, creative thinking is considered an important 21st century skill. Nevertheless, our understanding of how contextual factors such as socio-economic status (SES) and gender affect creativity is still limited – especially from an international perspective. In the current study, we thus examined the impact of gender and SES on creative thinking across 62 countries, using data from the 2022 Programme for International Student Assessment (PISA). Creative thinking, alongside mathematics and reading, was analyzed using a two-stage meta-analytic approach, integrating effect sizes from country-specific samples with a total sample size of $N = 493{,}660$. Our results revealed consistent gender disparities, with females outperforming males in creative thinking and reading, a trend robust across countries but with considerable variability. Gender disparities were less pronounced in the mathematical domain. Moreover, SES was found to be a strong predictor of creative thinking, mathematics, and reading, with higher SES associated with better performance across all domains. There was no substantial interaction effect between gender and SES for creative thinking and reading, suggesting that SES advantages are consistent across genders. Our analyses indicated substantial heterogeneity between countries, emphasizing the need for context-specific educational policies. These findings highlight the pervasive influence of gender and SES on fundamental educational outcomes and hence stress the necessity of tailored interventions to address these disparities around the globe.

Transparency and Openness: In this manuscript, we report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures relevant to the current study. We used data from the international 2022 PISA survey, which is publicly available from the OECD download pages as the source to build the database for the present study (https://www.oecd.org/en/data/datasets/pisa-2022-database.html). All data-analytic code to reproduce the current results is available via the Open Science Framework (https://osf.io/zbfg8/).

* Corresponding author.
E-mail address: academ@benjamin-goecke.de (B. Goecke).

https://doi.org/10.1016/j.tsc.2025.101963
Received 13 March 2025; Received in revised form 1 July 2025; Accepted 17 August 2025
Available online 18 August 2025

Educational Impact and Implications Statement

Creative thinking is considered a crucial 21st century and future workforce skill, but our understanding of how contextual factors like socio-economic status and gender affect this skill is still limited. Our current study, analyzing PISA 2022 data across 62 countries, revealed that female students outperform males in creative thinking tests and that students from higher socio-economic status backgrounds show better performance in these tests. As this finding was consistent across the globe, our study suggests a dire need to consider ways of how to reduce these apparent disparities, if students within countries and across the globe are to have the same opportunities of developing their creative thinking and hence their creative potential. The current findings thus imply the reconsideration of present educational policies. This will not only help in reducing educational inequalities but also in preparing students more effectively for the demands of the future workforce.

1. Introduction

Although creativity has been studied for almost a century, its relevance to education has increased, specifically in recent years (c.f. Graciano et al., 2023; Hernández-Torrano & Ibrayeva, 2020). The human capacity to think creatively, which involves coming up with original and useful ideas in a given social context (Plucker et al., 2004; Runco & Jaeger, 2012), is viewed as an essential skill for the future workforce in a world where humans are less reliant on physical work (Puccio, 2017) – even though advances in the field artificial intelligence continue to change our thinking about creativity (Ivcevic & Grandinetti, 2024; e.g., Rafner et al., 2023). Furthermore, creativity can be considered a crucial capacity for nations, as regional development appears to depend, among other factors, on novel combinations of knowledge and ideas (Mcgranahan & Wojan, 2007). Despite its importance on individual and group levels, our understanding of contextual factors that might facilitate or hinder creativity, such as socio-economic status (SES) and gender, is limited, especially from a perspective incorporating different cultures, countries, or economies (Acar et al., 2022; Taylor et al., 2024). Therefore, our study analyzes how SES and gender affected creative thinking, the innovative domain of PISA 2022, which was assessed in over 60 countries and economies. We compare the impact of these factors on creativity with their effects on mathematics and reading, which are fundamental competencies for academic achievement (e.g., Delgadova, 2015; Nguyen et al., 2016). This comparison helps to contextualize the challenges and opportunities in fostering creativity relative to more traditional academic domains like reading and mathematics.

1.1. Creativity assessment and education

Creativity is a multi-facetted construct, which has commonly been studied in terms of the ability to come up with diverse, creative ideas (Benedek et al., 2019; Forgeard & Kaufman, 2016), also known as divergent thinking (DT) ability (Guilford, 1967). DT ability can be measured with open-ended tasks that require the production of "alternative ideas, all of which satisfy a somewhat general requirement" (Guilford, 1977, p. 92). The current consensus among creativity researchers is that for DT assessments to be valid, instructions should require test-takers to generate creative responses (e.g., find creative uses of everyday objects; Diedrich et al., 2015; Harrington, 1975; Reiter-Palmon et al., 2019; Runco, 1991). DT performance can be scored in terms of the quantity (e.g., fluency, or flexibility) and quality (e.g., originality, creativity, or elaboration) of responses (Saretzki et al., 2024; Weiss et al., 2021). DT ability is considered an important indicator of human creative potential (Guilford, 1966; Lubart et al., 2010; Runco & Acar, 2012), which predicts individual differences in creative achievements together with general cognitive ability, personality and domain-specific expertise (Benedek, 2024; Said-Metwaly et al., 2020). Given the growing interest of creativity as a crucial characteristic in education, it has even been argued that the inclusion of creativity measures into high-stakes educational assessments, such as college admissions tests, could not only increase our understanding of student abilities by getting a more holistic idea about their potentials, but also help to reduce biases of college admission tests (Kaufman, 2010). Due to that, it seems only sensible that creative thinking was chosen as the innovative domain of PISA 2022. PISA's operationalization of creative thinking assessments aimed to assess central aspects of DT ability across different domains: Students were tasked with generating original or diverse ideas or with improving a given idea to stimuli for written or visual expression or responses to social or scientific problems (OECD, 2023a, 2024a).

1.2. Creativity in cross-cultural and comparative contexts

Studying creativity across cultures, countries or economies is complex due to varying concepts and methods (Barth & Stadtmann, 2024; Glăveanu, 2019). For meaningful comparisons, it is crucial that the definition of creativity is consistent across cultures and that the measurement instruments applied are invariant across contexts (e.g., He & Van De Vijver, 2012). As posited by Barth and Stadtmann (2024), cultures exhibiting greater individualism may be expected to demonstrate elevated levels of originality relative to those displaying lesser degrees of individualism. Conversely, cultures with stronger collectivist tendencies may be hypothesized to exhibit higher scores on appropriateness. It is evident that making comparisons across cultures for overall measures of creativity, which typically aim to combine both components of creativity (that is, appropriateness and originality) into one score, may be challenging when using these theoretically derived expectations. However, it is important to note that overall creativity scores are typically best explained by considering originality (Diedrich et al., 2015; Pichot et al., 2024). This suggests that comparing creativity

across cultures can still be valuable, at least to some extent. Indeed, Western cultures show a small to moderate advantage in creativity ($g = 0.33$, 95 %-CI: [0.17, 0.49]), according to a recent meta-analysis (Barth & Stadtmann, 2024). However, caution is warranted as only one of the 41 included studies tested for measurement invariance. Measurement invariance testing is typically absent in cross-cultural comparisons of creativity (for exceptions, see Guo et al., 2023, 2024; Kornilov et al., 2016), despite growing calls for its inclusion (Karwowski, 2016). Therefore, methods must be found that consider possible cultural differences, when comparing performances on different ability tests across cultures, countries, or economies.

Previous cross-cultural research identified differences in creativity across countries. For example, Niu and Sternberg (2001) found that in studies on artistic creativity, American participants received higher creativity and aesthetic ratings than Chinese participants; similar trends were observed in verbal creativity indicators using the Torrance Test of Creative Thinking were used (Jaquish & Ripple, 1985). However, there are issues with existing cross-cultural comparisons, although these studies offer a reasonable starting point for deriving expectations for further studies. For example, previous cross-cultural studies often suffer from relatively small participant samples (c.f.. Jellen & Urban, 1989), often drawn from a limited number of countries. More than that, some of the studied countries are more frequently sampled than others, which is problematic with respect to generalizability of results across the globe. As Niu and Kaufman (2013) described, previous studies have mostly relied on studying American and Chinese students. Although several other countries were considered in other studies (Corazza & de Saint-Laurent, 2020), many regions across the globe have been largely overlooked altogether so far. While recent efforts have expanded the geographic reach of creativity research (c.f. Barth & Stadtmann, 2024), a large-scale cross-country study that systematically investigates creative thinking on a global scale has been missing to date. For an overview of the current state of the literature and a comprehensive discussion of the limitations of cross-cultural research on creativity, we direct the interested reader to the work by Barth and Stadtmann (2024).

With our current work, we aim to overcome these limitations by analyzing larger and more diverse samples from various countries and economies, using creative thinking assessments designed for high conceptual and psychometric comparability.

1.3. Socio-economic status and creativity

Socio-economic status (SES) refers to the broad social standing of students and their families, typically capturing educational, occupational, and material resources, such as parental education levels, job status, or household possessions (e.g., Antonoplis, 2023; Avvisati, 2020; Ehmke & Siegle, 2005). The fact that SES has an influence on educational outcomes is well known (Broer et al., 2019). However, the relationship of SES with creativity is less clear, and not surprisingly contradictory hypotheses regarding this relationship exist. One might posit that there is a similar intertwining of SES and creativity as exists between SES and other educational domains such as reading and mathematics. That is, better situated persons could have advantages that are grounded in being blessed with more support and opportunities. An alternative argument would be that low SES fosters creativity due to limited resources that require creative problem-solving (Acar et al., 2022).

Recent meta-analytical evidence reported effect sizes from $k = 117$ studies and showed that SES and creativity are weakly related ($r = 0.12$, 95 % CI: [.09, 0.15]; Acar et al., 2022) indicating that people with higher SES tend to show higher scores on measures of creativity. This effect size varied slightly if only parental education ($r = 0.17$) or parental occupation ($r = 0.11$) were taken into account as measures for SES. In sum, previous evidence showed a small and positive effect of SES on creativity. However, this meta-analytical evidence points in the direction that the relationship of SES with creativity is less pronounced than the correlation of SES with other educational outcomes, academic achievement, and intelligence (e.g., mothers' education and academic achievement: $r = 0.33$; mothers' education and vocabulary: $r = 0.41$; Blums et al., 2017).

1.4. Gender and creativity

The study of gender differences in various cognitive abilities has been a prominent area of research, with significant implications for education and educational outcomes (Spinath et al., 2014). While many studies have documented gender disparities across different cognitive skills (for an overview, see for example Halpern, 2011), results concerning divergent thinking have been mixed and inconsistent. Abraham (2016) concluded that gender differences exist in general but vary across different areas (e.g., females tend to have more creative potential, but males tend to achieve more creative products). The author also provided insight into what factors might be driving these differences, such as the use of rewards or the consideration of different age groups. A recent meta-analysis of 213 studies indicates that females slightly outperform males in divergent thinking abilities ($g = -0.07$, 95 % CI [−0.10, −0.03]; Abdulla Alabbasi et al., 2022). This research also explored the hypothesis that males exhibit greater variability in divergent thinking, a pattern observed in other cognitive abilities (Hedges & Friedman, 1993). Indeed, male variability in divergent thinking was 21.6 % higher (lnVR = 1.22, 95 % CI [1.13, 0.1.29]; Abdulla Alabbasi et al., 2022). However, this finding was contested by another systematic review and meta-analysis, which investigated whether this greater male variability contributes to the underrepresentation of females among eminent creators (Taylor et al., 2024). The authors again reported a mean advantage for females in creative ability, mostly operationalized as DT ($g = -0.10$, 95 % CI [−0.13, −0.06]; Taylor et al., 2024), but only a negligibly higher variability in males (lnVR = 0.02, 95 % CI [.00, 0.04]; Taylor et al., 2024). Overall, these studies suggest a slight female advantage in divergent thinking, yet they deliver mixed conclusions on the hypothesis of greater male variability in creativity.

1.5. Socio-demographic disparities across academic domains

To interpret social disparities in creative thinking, it is useful to consider how SES and gender relate to academic achievement in

more traditional school subjects like reading and mathematics. These two domains are particularly relevant because creative thinking was shown to be associated with both (e.g., Bicer et al., 2021; Wang, 2012). This makes sense, as theoretically, creativity and reading may be linked through shared verbal and elaborative processes (e.g., Ritchie et al., 2013), while creativity and mathematics are connected via problem-solving and cognitive processes like pattern recognition (Bicer et al., 2021).

Beyond their theoretical links to creativity, reading and mathematics are also well-documented domains of educational stratification; that is, they consistently reveal performance gaps between social groups. For example, SES has been shown to be a predictor of student performance in reading and mathematics, with numerous studies and PISA cycles reporting that students from more advantaged backgrounds outperform their peers (e.g., Sirin, 2005), although the strength of the association can vary depending on the operationalization of SES and the educational context. Previous PISA cycles have all consistently revealed that social background is related to the competencies students show in mathematics, reading, and natural sciences (e.g., Müller & Ehmke, 2013), and the past pandemic of COVID-19 may have even exacerbated these social inequalities (Förster et al., 2023; Stanat et al., 2022). Previous PISA reports have revealed that, for example in Germany, SES explained 14.2 % of the variance in mathematical competencies, which is slightly above the OECD average of 11.7 % (Lewalter et al., 2023). In PISA 2022, however, there were no meaningful gender differences revealed for scientific competencies for 22 of the 37 OECD states, implying that in these countries a promotion of scientific skills has been relatively successful in reaching gender parity (Lewalter et al., 2023).

Similarly, large-scale assessments have repeatedly found that girls tend to outperform boys in reading across virtually all education systems (Brunner et al., 2023; Stoet & Geary, 2013). Brunner and colleagues (2023) have developed and used an integrative individual participant's meta-analytic approach to analyze the effects of gender, SES, and the interaction between both variables on reading achievement. Their findings revealed that female students, on average, outperformed male students with regards to reading performance across 92 countries ($\beta_{gender} = 0.37$; 95 %-CI [.35, 0.39]). Similarly, they showed that an increase of 1 *SD* in SES (as indicated by HISEI) was associated with an increase of 0.30 *SD* in reading achievement ($\beta_{SES} = 0.30$; 95 %-CI [.29, 0.31]). Interestingly, however, the interaction of gender and SES hardly had an impact on reading achievement ($\beta_{genderxSES} = -0.02$; 95 % CI [−0.02, −0.01]), that is, there was no moderating effect. In mathematics, however, the gender pattern seems more complex: while average differences are generally small, boys tend to outperform girls at the upper end of the performance distribution, particularly in higher-achieving countries, whereas in some countries girls perform equally well or even better (Stoet & Geary, 2013).

Drawing on reading and mathematics as comparative academic domains is therefore based on a dual rationale: not only are they established indicators of educational stratification by gender and SES, but they also correlate with creative thinking in theoretically and empirically meaningful ways. This allows us to explore whether the social disparities found in traditional academic outcomes also manifest in students' creative thinking.

1.6. Potential macro-level moderators of performance disparities

Cross-national variability in educational outcomes is not solely a function of individual characteristics but may also reflect broader societal structures on the country-level. Several such macro-level indicators have been proposed to explain how gender and SES disparities in education manifest differently across countries (e.g., Stoet & Geary, 2013). While the following list is not exhaustive, the indicators represent theoretically and empirically grounded constructs that are particularly relevant to the current investigation and have been used in previous comparative education research (Brunner et al., 2023; Else-Quest et al., 2010; Guiso et al., 2008; Reilly, 2012; Stoet & Geary, 2013).

First, membership in the Organization for Economic Co-operation and Development (OECD) may serve as a meaningful moderator variable for socio-demographic disparities in academic outcomes. Not only has the OECD pledged to tackle gender gaps (OECD, 2012; Closing the Gender Gap: Act Now)—which might be observable in comparison between OECD and non-OECD countries (cf., Barra & Boccia, 2022)—, but member states also tend to share institutional features such as more standardized curricula and greater investment in educational infrastructure, all of which may interact with socio-demographic variables like gender or SES. Moreover, because OECD membership largely overlaps with Western cultural contexts, variations in gender role socialization and value systems may further shape the manifestation of gender disparities. This is also related to the next potential moderator variables.

Second, gender equality, as measured by the Global Gender Gap Index (GGGI), reflects the extent to which societies provide equitable opportunities for women and men across education, politics, economy, and health (cf., The Global Gender Gap Report 2022, 2023). In more gender-equal societies, gender gaps in education may be smaller due to diminished cultural or institutional barriers—something that could, for example, be observed in reading where girls typically outperform boys overall (Eriksson et al., 2020). Conversely, persistent gaps may signal deeper structural inequities that extend into schooling and adolescence, like in the context of PISA.

Third, more broadly, a country's Human Development Index (HID; e.g., United Nations Development Programme, 2025), which captures national averages in health, education, and income, may moderate educational inequalities. While higher HDI generally implies broader access to educational resources, prior research suggests that high-HDI societies may also foster greater individualism, thereby enabling the expression of traditional gender-typed academic preferences (Charles & Bradley, 2009), like for example fewer women in STEM-related fields (e.g., Cheryan et al., 2024). For SES effects, higher HDI may be associated with more developed educational infrastructures and broader access to schooling; however, it may also coincide with institutional practices such as early tracking or selective admissions that can amplify already existing achievement gaps.

Fourth, power distance, a cultural dimension proposed by Hofstede, captures the degree to which "the less powerful member of institutions and organizations within a country expect and accept that power is distributed unequally" (Hofstede et al., 2010, p. 61). In high power distance societies, people are more likely to accept social hierarchies as normal or justified. This can make inequalities

based on SES or gender more pronounced, leading to larger gaps in student achievement. Schools in these contexts often stress obedience and respect for authority, which may limit students' independence and make it harder for those from disadvantaged backgrounds to succeed and creativity to unfold (Huber & Paule-Paludkiewicz, 2024; Javidan et al., 2006).

Together, these indicators offer a valuable lens through which to examine how national contexts shape the interplay of gender, SES, and educational outcomes—particularly in the domain of creative thinking.

1.7. The present study

In the present study, we analyze data from the Programme for International Student Assessment (PISA) 2022 regarding the innovative domain (creative thinking) and established core domains (reading and mathematics). The domain of science was included as an auxiliary variable in the imputation process of SES (see methods section). We aimed to answer the following questions:

1. Is there a female advantage in creative thinking? If yes, is this advantage comparable across countries? Are gender differences similar to the differences reported in established domains like mathematics and reading?
2. Does SES predict creative thinking? Is this effect comparable across countries? Is SES equally influential for creative thinking performance as it is for established core domains of mathematics and reading?
3. Are there interaction effects of gender and SES on creative thinking?
4. Are the country-level effects moderated by macro-level variables like OECD membership, Global Gender Gap Index, Human Development Index, or the cultural dimension of power distance?

To answer these questions, we applied an integrative meta-analytical approach of individual participant data, as this analysis offers the opportunity to strengthen empirical evidence of consistency and generalizability (Brunner et al., 2023). This analysis is specifically tailored to complex survey designs (e.g., sampling weights, clustered and missing IPs) and allows the conduction of descriptive analysis, but also the integration of the results based on meta-analytical models (Brunner et al., 2023). The research objectives of the present study were not preregistered.

2. Method

2.1. Database (Participants and procedure)

We used a subset of data from the 2022 PISA survey, which is publicly available from the OECD (https://www.oecd.org/en/data/datasets/pisa-2022-database.html). As this study involves secondary data analysis, no additional ethical approval was required. Approximately 650,000 students from 81 countries and economies participated in PISA 2022, with each country and economy providing a representative sample of more than 6300 students, roughly balanced by gender. Students were assessed in mathematics, reading, and science, as well as creative thinking, using a matrix-sampling design (Rutkowski et al., 2013), where both students and items were drawn representatively. Following the cognitive tests, students, along with their parents, principals, and teachers, completed questionnaires to assess the context of their learning environment. Our subset included 62 of the 64 samples from countries or economies that participated in the innovative domain of PISA 2022 (creative thinking). In our dataset (downloaded on 25th July 2024), no creative thinking data were available for Cyprus. Additionally, we excluded the sample from Costa Rica from our analysis because the SES variables were not available. Consequently, our analysis included data from $N = 493{,}660$ students, with a mean age of 15.79 years ($SD = 0.06$), across 62 countries and economies globally. The descriptive overview of the available data for each included country or economy can be found in the supplementary materials (Table SM 1).

2.2. Measures

For examining our research questions, we focused on five variables from the available data: gender, SES as indicated by the index for economic, social and cultural status (ESCS), mathematics competence, reading competence, and creative thinking competence.

Gender. Gender was obtained from school records and validated by comparing it to the student responses (OECD, 2024b). We recoded the data so that female students were coded as 0 and male students were coded as 1 (i.e., female students are the reference group).

Socio-Economic Status. We used the Economic, Social and Cultural Status (ESCS) index to operationalize the SES of students. ESCS is a standardized composite measure of SES, integrating information on parental education, parental occupation, and household possessions (e.g., books at home). It thus captures multiple dimensions of students' socio-economic background (Akukwe & Schroeders, 2016), including economic, educational, and cultural resources. In the OECD PISA dataset, ESCS is scaled to have an international mean of zero and a standard deviation of one, enabling meaningful cross-country comparisons. For robustness, we also conducted the main analyses using the Highest International Socio-Economic Index of Occupational Status (HISEI; e.g., Avvisati, 2020), another widely used SES proxy based on the highest parental occupational status. For the sake of briefness, more information regarding the HISEI and the corresponding results can be found in the supplementary materials—the results were highly consistent with the findings based on ESCS as presented in the main manuscript, suggesting robust findings across SES indicators.

Competence in Cognitive Domains. Four cognitive domains were assessed in PISA 2022: mathematics, reading, science, and creative thinking. The measurement instruments of these cognitive domains were constructed to assess students' literacy. Literacy is defined as

the competence to apply in everyday situations the knowledge students accumulated in any given domain until they are 15-years old. Fifteen was chosen as the target since at that age compulsory schooling ends in most countries. The cognitive domains were understood as follows:

1. Creative Thinking covered three ideation processes: generating diverse ideas, generating creative ideas, and evaluating and improving ideas. Generating diverse ideas requires the student's ability to be flexible in idea generation. Generating creative ideas mostly assessed the ability to think outside the box and find original ideas (e.g., comparable to statistical infrequency of the given response category). Evaluating and improving ideas focusses on the ability to refine previously generated ideas. These ideational processes were situated in four context domains (written expression, visual expression, social problem solving, and scientific problem solving), with a total of 32 items. For sample items see OECD (2023a). One example in the facet of generative ideas is the item "*Illustration Titles*: Write 3 different titles for the illustration on the right. The titles should be as different from each other as possible". Human coders evaluated whether responses were appropriate and creative. To do so, a partial credit coding was applied, that is: coders first evaluated the appropriateness and would assign no credit (0) for inappropriate responses; for appropriate responses, partial credit (1) was assigned for unoriginal ideas, and full credit (2) for original responses. After this procedure, only one total creative thinking score was computed across ideation processes and context domains. For instance, if a student's response was appropriate but lacked creativity, it would only receive partial credit (OECD, 2024a). Full credit was awarded when students produced responses that were also considered original – for instance, including a novel combination of elements, such as lines, shapes, or colors—making responses stand out and unique (Cropley, 1967; Wilson et al., 1953). For those items where students had to provide multiple different ideas, human coders also checked whether these ideas were different from one another – this time, in the sense of flexibility (c.f. Reiter-Palmon et al., 2019). This assessment looked at whether the participants were able to provide different kinds of ideas, rather than simply reuse variations of the same concept. Vivid examples of different responses and subsequent scores can be found in the report of the OECD (2024a; pp. 129–130).
2. Mathematics competence in PISA is defined as "an individual's capacity to reason mathematically and to formulate, employ, and interpret mathematics to solve problems in a variety of real-world contexts. It includes concepts, procedures, facts and tools to describe, explain and predict phenomena. It assists individuals to know the role that mathematics plays in the world and to make the well-founded judgements and decisions needed by constructive, engaged and reflective 21st Century citizens" (OECD, 2023b, p. 22). Mathematics was the major domain in PISA 2022 and thus was assigned more test time/student responses than the other domains. Additionally, the theoretical framework and items are renewed for the major domain, whereas the minor domains are assessed based on the material from the year it was the major domain for the last time (i.e. 2018 for reading and 2015 for science; Diedrich & Lewalter, 2023).
3. Reading competence in PISA is defined as "understanding, using, evaluating, reflecting on and engaging with texts in society" (OECD, 2019, p. 28). This definition implies that reading competence is understood as a basic competence important for societal participation. The tasks covering this domain can be described by different dimensions: Source (single vs. multiple source), text format (continuous and non-continuous texts), organization (static vs. dynamic), and text type (e.g., description, narration, argumentation). All these dimensions can be fitted into a situational framework (personal contexts, public contexts, educational and occupational contexts). The important sub-skills that are assessed based on these tasks and domains are reading fluency, locating information, understanding texts, and evaluating and reflecting (Heine et al., 2023).

Importantly, in the PISA study, plausible values (PVs) were used to estimate students' competence across domains, ensuring more reliable group-level statistics. Therefore, ten plausible values per domain are usually estimated based on the population model (von Davier et al., 2009). To this end, all data of any given student—both responses to all cognitive domains and to the student questionnaire—are taken into account in a combined Item Response Theory (IRT)- and latent regression model (OECD, 2024a). Reading, mathematics, and science are reported on a scale with a range from 0 to (theoretically) over 800 points, split into 6 competency levels (see Chapter 14 and 17, OECD, 2024b). In the first year a domain was the major domain, the scales were calibrated to a mean of 500 and a standard deviation of 100 (OECD, 2024b). Unlike mathematics, reading, and science, scores for creative thinking were summarized on a bounded unidimensional scale ranging from 0 to 60 points (for more information see OECD, 2024b).

Moderator Variables. We used four moderator variables to examine whether macro-level indicators could explain between-country variation in socio-demographic disparities across creative thinking, mathematics, and reading. These variables were included in meta-regression models.

1. OECD Membership status. We included a binary indicator denoting whether a country or economy was a member of the Organisation for Economic Co-operation and Development (OECD). In total, $n = 26$ countries of the current subset of data were members of the OECD.
2. The Human Development Index (HDI) was used as a proxy for general socio-economic development. It captures national averages in health, education, and income. HDI scores were included as a continuous variable (0–1) and were obtained from the UNDP database (https://hdr.undp.org/data-center/human-development-index#/indicies/HDI). In total, the HDI was available for $n = 55$ countries.
3. The Global Gender Gap Index (GGGI; The Global Gender Gap Report 2022, 2023) captures gender equality in four key dimensions: economic participation, educational attainment, political empowerment, and health and survival. This index was included as a continuous variable (0–1). The respective data were taken from the Global Gender Gap Report 2022 (The Global Gender Gap Report 2022, 2023). In total, the GGGI was available for $n = 57$ countries.

4. Power Distance (PD) reflects the extent to which less powerful members of a society accept and expect power to be distributed unequally (Hofstede et al., 2010). Higher values indicate greater societal acceptance of hierarchical structures. The data were taken from the Hofstede Dimension Data Matrix in its most recent version (2015; https://geerthofstede.com/research-and-vsm/dimension-data-matrix/). In total, the power distance index was available for $n = 44$ countries.

3. Data preparation and treatment

3.1. Statistical analysis

We used a two-stage meta-analytic integrative method to analyze the available PISA 2022 assessment data (Brunner et al., 2023). In a first step, effect sizes are computed for all independent samples (i.e., countries or economies). In a second step, these effect sizes are meta-analyzed, similar to how traditional meta-analyses synthesize findings from independent studies or datasets. Therefore, this approach, which is equivalent to a two-step individual participant meta-analysis, has two advantages: first it considers the independent sampling of the PISA data within the respective countries. Second, the differential effects of the models between the respective countries can be analyzed more granularly. In the following, we will outline our analytical strategy. Please note that all scripts are available in an open science repository: https://osf.io/zbfg8/.

Data Preparation. Notably, there was a very small amount of missing data for the binary gender variable, with $n = 74$ observations (< 0.015 % of the total sample) removed through listwise deletion. We considered listwise deletion of this amount of missingness appropriate, aligning with previous work (Brunner et al., 2023; Graham, 2009). Additionally, we observed missing values for PVs on mathematics, reading, and creative thinking in Iceland, where $n = 4$ rows (< 0.12 % of the sample of Iceland) contained missing data. These rows were also removed via listwise deletion, as the completeness of all other PVs across all other domains and countries was given, and we could not rule out that these four rows were corrupted in the original data.

Multiple Imputation Process. On average, 4.51 % of cells were missing for the ESCS variable per country within the dataset. As missing at random could be assumed, we imputed the missing data applying a nested multiple imputation procedure using a Bayesian approach with non-informative prior distribution, with all variables standardized prior to imputation to match the scale used in the analysis models, similar to Brunner and colleagues (2023) adopting a substantive-model-compatible modeling approach (Enders et al., 2020; Grund et al., 2021). To this end, we performed $M = 10$ imputations for each of the independent 62 student sample data sets separately for each respective PV sets. The imputation model comprised of two conditional models. In the outcome model, we regressed the respective PV of creative thinking on gender, the respective variable of SES, their interaction and a random intercept for schools (to account for the clustered data structure with students being nested in schools). In the covariate model, we regressed the respective SES variable on gender and the respective PV of science competence as an auxiliary variable. Again, we specified a random intercept for schools. We used science competence as an auxiliary variable in this process due to its consistent relationships with students' SES and gender across countries (Sirin, 2005). Including auxiliary variables in an imputation model is recommended in the literature because it can strengthen the plausibility of the missing-at-random assumption and mitigate potential biases in parameter estimates (Lüdtke et al., 2020). In our case, using science competence as an auxiliary variable captures variance helps account for shared variance with SES, thus producing more accurate imputations and reducing potential bias in the later estimated association between SES and creative thinking. Due to the applied approach, the imputation models were constructed to be compatible with the subsequent regressions analysis for effect size estimation. We ran this procedure for 6.500 iterations (HISEI) and for 7.500 iterations (ESCS). We used the R package mdmb for the imputation process (Robitzsch & Lüdtke, 2024).

Effect Size Estimation and Sampling Variances. We specified and estimated regression models for each of the independent country datasets for each PV in each of the three target domains (creative thinking, mathematics, and reading) to obtain effect sizes and sampling variances. In total, we thus ran $62 \times 10 \times 3 = 1860$ regression models. To account for the complex sampling design used in PISA, we conducted weighted regression analyses that consider the hierarchical structure of the data, with students nested within schools. We applied a design-based estimation method that incorporates both student-level sampling weights and replicate weights, following the Balanced Repeated Replication (BRR) technique with Fay's adjustment (Rutkowski et al., 2010). This approach provides robust standard errors and allows for accurate population-level inferences. All models were estimated separately for each country and each imputed dataset, and the results were combined using Rubin's rules (Rubin, 1976). This strategy aligns with current best practices for analyzing large-scale educational assessment data (Brunner et al., 2023).

In the first regression model, we predicted the *z*-standardized PVs by the non-standardized gender variable (0 = female, 1 = male). In this model, the standardized regression coefficient can be interpreted as the standardized mean difference (equivalent to Cohen's *d*) between male and female students in the respective outcome PV, with negative values indicating that female students outperformed male students.

In the second regression model, we predicted the *z*-standardized PVs by the *z*-standardized SES variable. Here, the standardized regression coefficient can be interpreted as the correlation between the SES variable and the respective outcome PV, reflecting how much the outcome increases (in terms of standard deviations) for each one standard deviation increase in SES.

In the third regression model, we predicted the z-standardized PVs by the non-standardized gender variable, the *z*-standardized SES variable, and an interaction term between gender and SES. The interpretation of the standardized regression coefficients of the main effects is analogous to the first two models: the coefficient for gender reflects the standardized mean difference between male and female students, and the coefficient for SES reflects the correlation between SES and the outcome. The standardized regression coefficient of the interaction term, however, can be interpreted as indicating how gender differences in the respective outcome PV (in terms of Cohen's *d*) change when students' SES increases by 1 standard deviation.

We then computed pooled effect sizes and sampling variances across all imputations for each country using Rubin's rules. This pooling process involved calculating the average effect size (e.g., β_{Gender}, β_{SES}, $\beta_{Gender\times SES}$) across the imputed datasets. The pooled effect sizes and sampling variances were computed separately for each country and each PV. This approach ensures that the final estimates are robust. We then meta-analyzed these average standardized regression coefficients as effect sizes (Nieminen, 2022).

Additionally, we computed gender-specific, survey-weighted variances of the creative thinking PVs for each dataset and expressed the female–male dispersion contrast as a natural log variance ratio (lnVR), following Taylor and colleagues (2024). Positive values indicate greater variability among female students. These lnVR estimates were pooled using Rubin's rules and subsequently meta-analyzed. This approach allowed us to extend our analysis beyond mean-level differences and formally assess potential gender differences in variability.

Meta-Analytical Approach. We applied a cluster-robust random-effects model with a restricted maximum likelihood estimator (Pustejovsky & Tipton, 2022; Viechtbauer, 2010) in a traditional two-level model with students at level 1 and the independent studies (countries/economies) at level 2. All models were estimated using the R package *metafor* (Viechtbauer, 2010) and *clubSandwich* (Pustejovsky & Tipton, 2022). For our main analysis, we focused on the standardized regression coefficients computed in regression model 3 (c.f., β_{Gender}, β_{SES}, $\beta_{Gender\times SES}$). We ran a meta-analytic model per type of standardized regression coefficient and per competence domain. That is, we ran 9 meta-analytical models. For each model, we meta-analyzed $k = 62$ independent effect sizes, corresponding to the number of individual countries or economies in our subset of data. Model estimates for the univariate regression models (gender or SES as sole predictors) can be found in the supplementary materials (SM Table 2; for HISEI, cf. SM Table 8). In addition to that, we report three-level meta-analytical models with dummy-coded domains (creative thinking as the reference domain; that is, regression effect sizes are to be interpreted relative to this domain) in the supplementary materials (SM Tables 3–5; for HISEI, cf. SM Tables 9–11), to depict differences between competence domains in single models.

Heterogeneity between countries in the observed effect sizes was quantified by the I^2-statistic (Higgins & Thompson, 2002), which describes the proportion of between-country variation in effect size estimates that is due to true heterogeneity (i.e., the proportion of heterogeneity that cannot be explained by sampling error). Following the often-applied rules of thumb, I^2 of 0.25, 0.50., and 0.75 indicate low, medium, and high heterogeneity, respectively (Higgins et al., 2003). Additionally, we present τ^2 estimates as an indicator of the variance of the true effect sizes between countries (Viechtbauer, 2010).

We examined moderation effects on the pooled effect size using mixed-effect regression analysis (Viechtbauer, 2010). To investigate moderator effects, we adhered to the same estimation technique as described above but used meta-regression models by adding a predictor variable to the estimation of the effect sizes. Again, we specified 9 meta-regressions (for each of the standardized regressions coefficients and per competence domain).

We present 95 % confidence intervals, 95 % prediction intervals, and interpret *p*-values below 0.05 as significant. We visualized our findings by using an adaptation of forest plots (i.e., a graphical display of the observed effects and their corresponding confidence intervals).

Table 1
Effects of full model (Outcome = Gender + SES + Gender*SES).

				95 %-CI		95 %-PI				
Predictor	Outcome	Est. (SE)	*p*	*LB*	UB	*LB*	UB	Q (df = 61)	τ^2	I^2
Gender	Crea	−0.277 (0.014)	< 0.001	−0.304	−0.250	−0.472	−0.081	428.80 ($p < .001$)	.009	87.43
	Math	.028 (0.015)	.074	−0.003	.058	−0.198	.253	872.82 ($p < .001$)	.012	93.68
	Read	−0.294 (0.014)	< 0.001	−0.322	−0.256	−0.501	−0.087	633.56 ($p < .001$)	.011	92.07
SES	Crea	.334 (0.009)	< 0.001	.315	.353	.192	.476	606.64 ($p < .001$)	.005	89.52
	Math	.361 (0.010)	< 0.001	.342	.381	.213	.510	922.71 ($p < .001$)	.005	92.79
	Read	.342 (0.009)	< 0.001	.324	.361	.205	.479	684.50 ($p < .001$)	.005	90.78
Gender*SES	Crea	−0.034 (0.005)	< 0.001	−0.044	−0.023	−0.088	.021	113.85 ($p < .001$)	.000	46.84
	Math	−0.006 (0.006)	< 0.001	−0.018	.006	−0.066	.054	132.09 ($p < .001$)	.001	54.45
	Read	−0.026 (0.006)	< 0.001	−0.037	−0.015	−0.085	.033	126.28 ($p < .001$)	.001	52.77

Notes. Crea = Creative Thinking, Math = Mathematics, Read = Reading, *SE* = Standard Error, 95 %-CI = 95 % Confidence Interval, 95 %-PI = 95 % Prediction Interval, LB = Lower Boundary, UB = Upper Boundary, *Q* = Homogeneity Statistic, τ^2 = Variance of True Effects, I^2 = percentage of total variation in effects across countries that is due to true between-country heterogeneity rather than chance. Gender coding: 0 = female, 1 = male.

4. Results

4.1. Meta-analysis of the relationship between cognitive domains, gender, and SES

Gender. The overall estimates for creative thinking, mathematics, and reading from the cluster-robust random-effects meta-analysis can be found in Table 1 (for similar results regarding the HISEI, cf. SM Table 6).

Our findings reveal significant gender differences in all three domains. Female students outperformed male students in creative thinking tasks, as indicated by a significant negative estimate for gender. This effect is robust across countries with a true between-country variance of $\tau^2 = 0.009$, indicating moderate heterogeneity in the size of this main effect. Thereby, the largest effect sizes could be observed for Jordan (JOR), the Palestinian territories (PSE), and Saudi-Arabia (SAU). The smallest effect sizes were found for Peru (PER), Chile (CHL), and Mexico (MEX). In mathematics, the gender differences were not statistically substantial, but the τ^2 value of 0.012 also suggests variability across countries. In reading, female students again demonstrated stronger performance compared to male students, and this effect was consistent, though with some variability across countries as indicated by an τ^2 value of 0.011. An overview of effect sizes for each country or economy within each domain is depicted in Fig. 1 (for HISEI, cf. SM Fig. 3). In addition to that, we have depicted these effects on a global map (Fig. 2; for HISEI, cf. SM Fig. 4) to give a spatial impression of the global distribution of countries that were part of the current analysis.

To complement the mean-level analyses, we also examined gender differences in the variability of creative thinking performance across countries using the above-mentioned lnVR estimates. This approach allowed us to assess whether the spread of creative thinking scores differs between female and male students, beyond average performance gaps. A cluster-robust random-effects meta-analysis across the 62 participating countries and economies resulted in a pooled lnVR of −0.032 ($SE = 0.013$, $p < .05$), suggesting that female students' creative thinking scores were, on average, slightly less variable than those of their male peers. Back-transforming this estimate yields a variance ratio of 0.97 (95 %-CI: [.94 – 0.99]; 95 %-PI: [−0.24 - 0.18]), indicating approximately 3 % lower score dispersion among females. While statistically significant, this effect is modest in magnitude. Importantly, heterogeneity in these lnVR estimates was considerable, as reflected by a τ^2 value of 0.105, indicating variation in gender-based variability ratios across countries

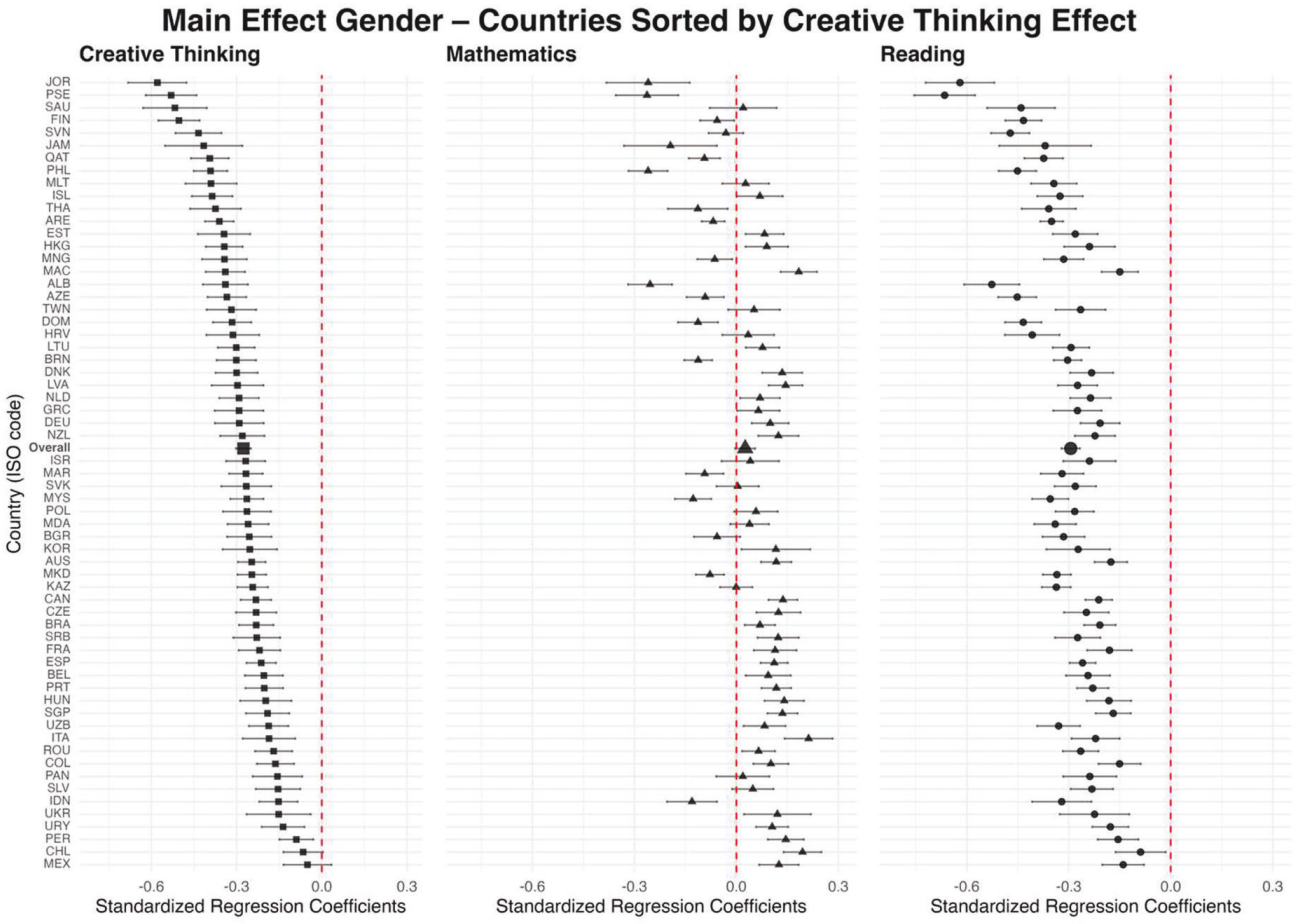


Fig. 1. Forest plot of the main effect of gender (Controlled for SES and Gender*SES).
Notes. Countries are denominated regarding their ISO-3166–1 A-3 Codes (https://www.iso.org/iso3166-country-codes.html). The overall meta-analytical effect size for each domain is presented at the top of each panel. The depicted intervals for the single effects are 95 % confidence intervals. A positive effect sizes indicates advantages for males, whereas a negative effect size indicates advantages for females.

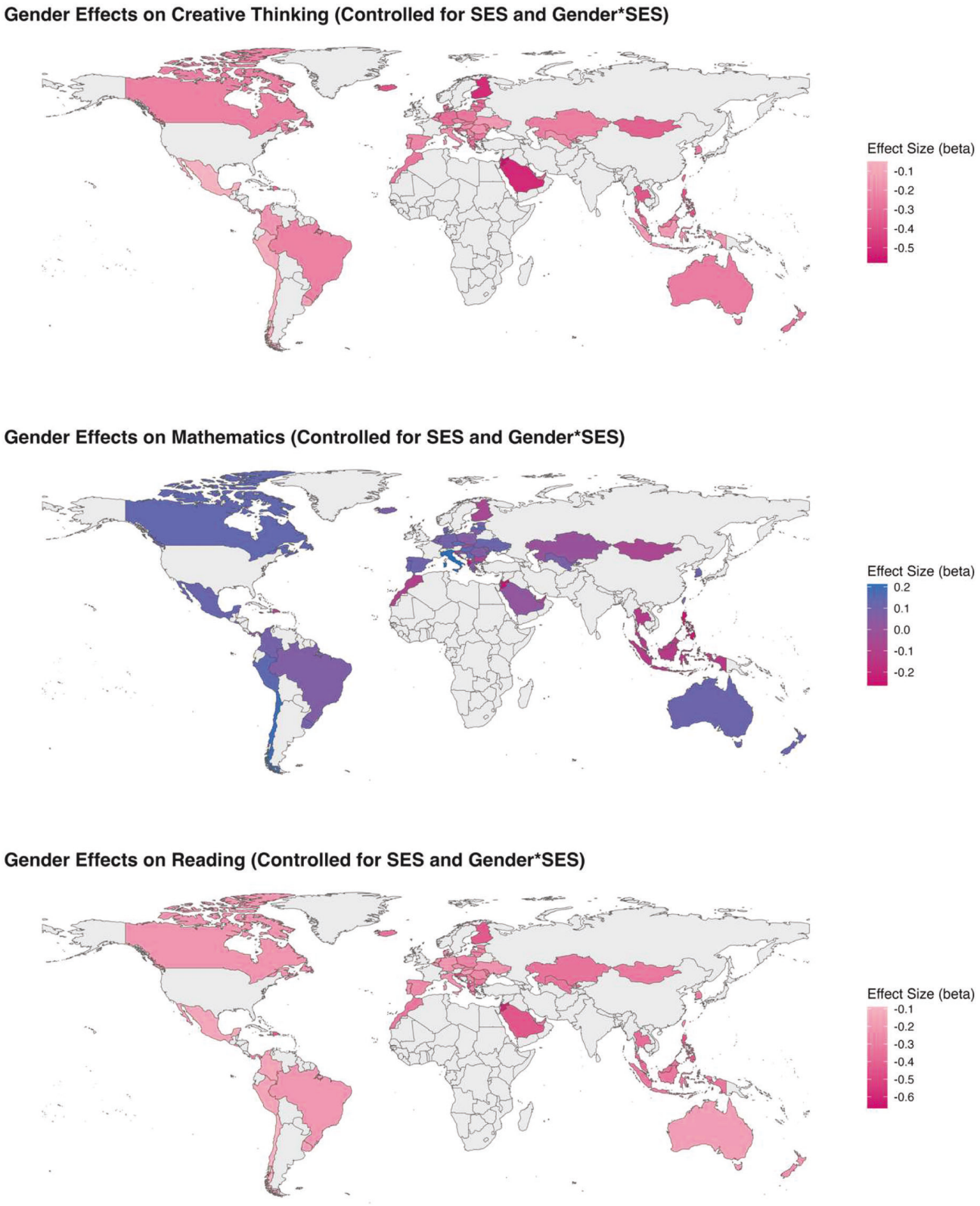


Fig. 2. Map of the main effect of gender (Controlled for SES and Gender*SES).
Note. A positive effect sizes indicates advantages for males, whereas a negative effect size indicates advantages for females.

and economies.

SES. SES was also observed as a strong and consistent predictor of performance across all three domains. Higher SES was associated with better performance in creative thinking, mathematics, and reading. The effect sizes suggest that for each unit increase in SES,

there was a corresponding increase in performance by approximately 0.33 *SD*s in creative thinking, 0.36 in mathematics, and 0.34 in reading. True between-country variance in this effect was reflected in τ^2 estimates of 0.005 for all three domains. Regarding creative thinking, the smallest effects were observed for Korea (KOR), Uzbekistan (UZB), and Jamica (JAM), whereas the largest effects were found for Bulgaria (BGR), Hungary (HUN), and Romania (ROU). Again, the effect sizes for each country or economy within each domain are depicted in Fig. 3 (for HISEI, cf. SM Fig. 5), and an additional map is provided in the supplementary materials (SM Fig. 1; for HISEI, cf. SM Fig. 6).

Gender x SES Interaction. While the interaction between gender and SES was statistically significant across all three domains, the effect size was smallest for mathematics, followed by reading, and largest for creative thinking. For creative thinking, the negative and significant interaction suggests that the SES advantage is slightly larger for female students than for male students. However, the estimated τ^2 value of 0.000 indicates very low heterogeneity, suggesting that this pattern is highly consistent across countries. In mathematics, the interaction effect was also negative, indicating that the SES advantage may be somewhat stronger for female students – the corresponding τ^2 value of 0.001 suggests limited variation across countries. Similarly, in reading, the interaction effect was small and negative, with $\tau^2 = 0.001$, again suggesting a largely consistent cross-national pattern. These results imply that across domains, there is a very slight trend for high SES to benefit female students more than male students, though the effects are rather small. The full set of interaction effects is illustrated in Fig. 4 (for HISEI, cf. SM Fig. 7), with an additional map provided in the supplementary materials (SM Fig. 2; for HISEI, cf. SM Fig. 8).

4.2. *The moderating effect of macro-level variables*

To better understand the contextual variation in disparities across countries, we examined whether macro-level variables of these respective countries moderate the observed effects of gender, SES, and their interaction on the student performance in creative thinking, mathematics, and reading. Specifically, we considered four macro-level moderator variables: OECD membership, the Global Gender Gap Index (GGGI), the Human Development Index (HDI), and Power Distance (PD). The corresponding results can be found in Table 2 (for HISEI and OECD membership, cf. SM Table 7).

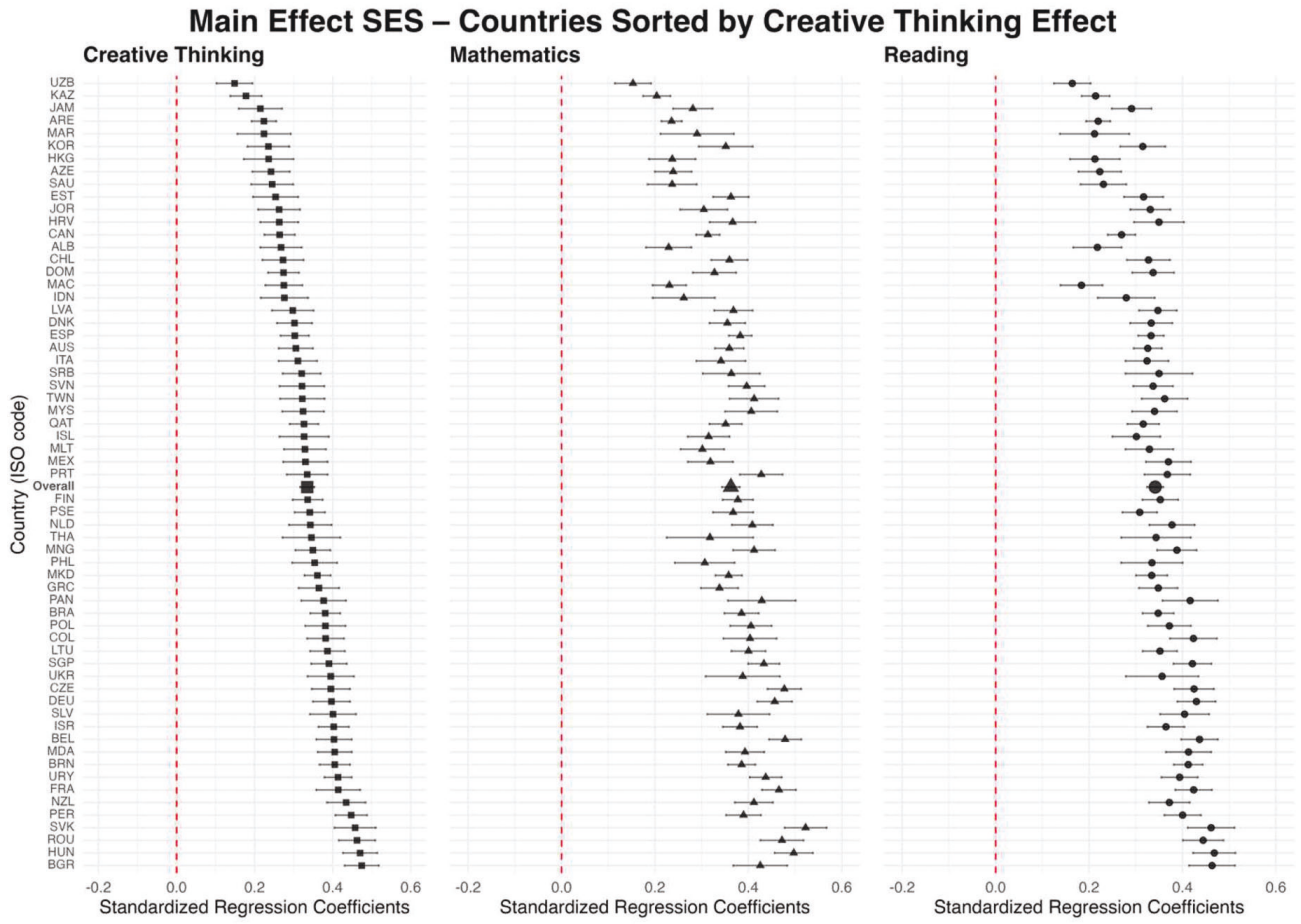


Fig. 3. Forest plot of the main effect of SES (Controlled for Gender and Gender*SES).
Notes. Countries are denominated regarding their ISO-3166–1 A-3 Codes (https://www.iso.org/iso3166-country-codes.html). The overall meta-analytical effect size for each domain is presented at the top of each panel. The depicted intervals for the single effects are 95 % confidence intervals. A positive effect sizes indicates advantages for males, whereas a negative effect size indicates advantages for females.

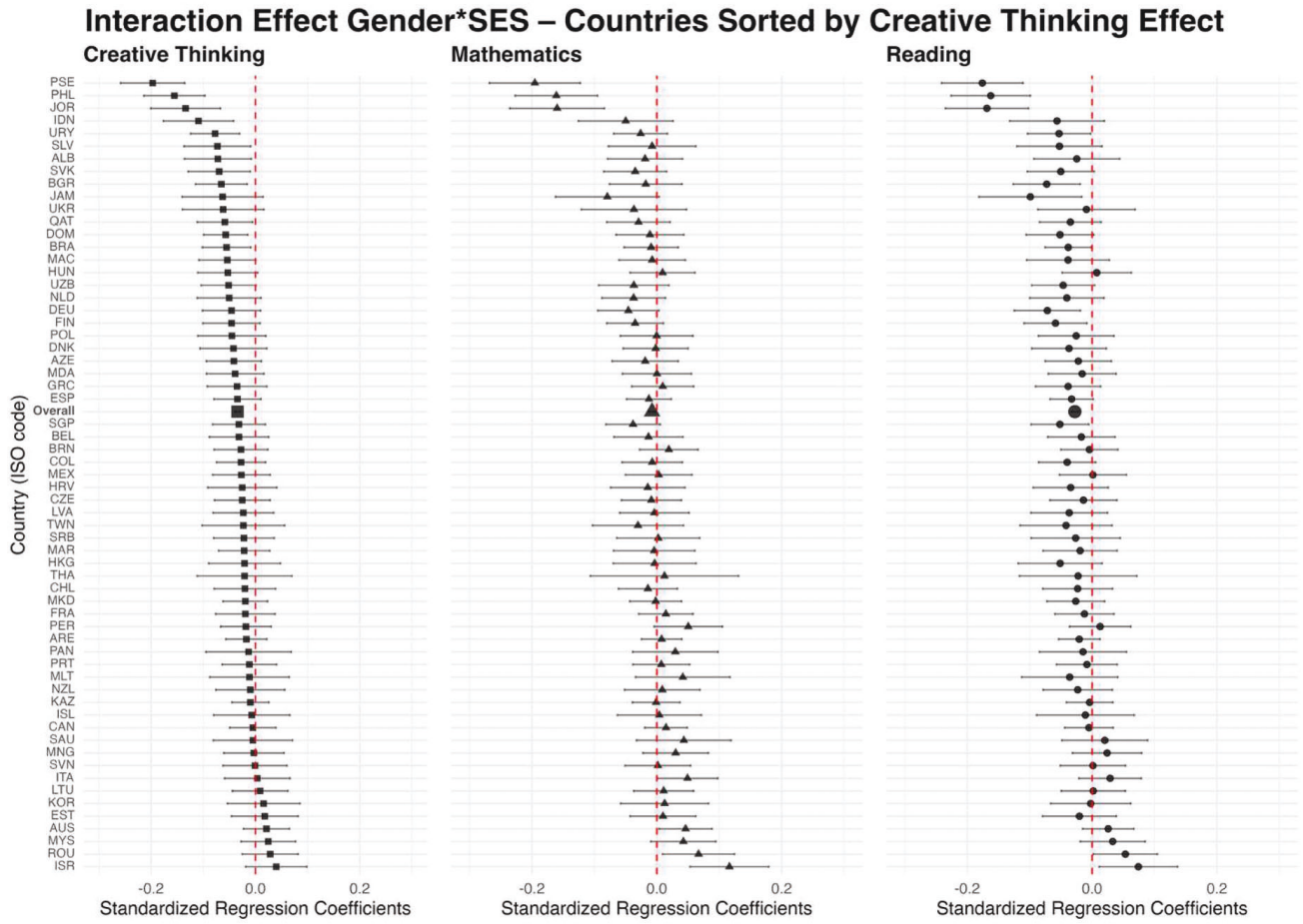


Fig. 4. Forest plot of the interaction effect of gender*SES (Controlled for Gender and Gender*SES).
Notes. Countries are denominated regarding their ISO-3166–1 A-3 Codes (https://www.iso.org/iso3166-country-codes.html). The overall meta-analytical effect size for each domain is presented at the top of each panel. The depicted intervals for the single effects are 95 % confidence intervals. A positive effect sizes indicates advantages for males, whereas a negative effect size indicates advantages for females.

For creative thinking, none of the macro-level variables significantly moderated the main effects of gender or SES, suggesting that these effects are highly stable across diverse national contexts, regardless of economic development, gender equity, or cultural hierarchy. However, one statistically significant moderator for the interaction between gender and SES could be observed: OECD membership. The interaction effect was slightly stronger in OECD countries, indicating that in these countries, the SES advantage may be more similar for boys and girls, while in non-OECD countries, the SES advantage might lean slightly more toward one gender. Although the effect was small, the consistency across countries suggests a stable pattern.

For mathematics, several macro-level variables significantly moderated the observed effects. The gender main effect was more pronounced in OECD countries and in countries with a higher HDI, suggesting that economic and societal development may be associated with stronger gender differences. In contrast, gender differences were smaller in countries with higher power distance, suggesting that more hierarchical societies may have smaller gender gaps. Students' socio-economic background also had a stronger effect on mathematics performance in OECD countries, possibly because more developed education systems can increase SES-related differences. We did not find any significant effects for the interaction between gender and SES in this domain.

In reading, both OECD membership and HDI significantly moderated gender effects, with female students showing greater advantages in more developed and economically advanced countries. Similarly, SES effects on reading were more pronounced in OECD countries. Importantly, the interaction between gender and SES was also moderated by HDI. No other significant moderator effects were found for reading.

Lastly, moderation analyses examining the heterogeneity in gender-based variability in creative thinking (as reflected by lnVR effect sizes) revealed several significant macro-level predictors. OECD membership was associated with reduced gender variability, with lower male-to-female score dispersion in member countries ($b = -0.095$, 95 %-CI [−0.141, −0.050], $p < .001$; 95 %-PI: [−0.19 - 0.20]). Similarly, higher HDI scores predicted significantly smaller gender variance gaps ($b = -0.909$, 95 %-CI [−1.22, −0.598], $p < .001$; 95 %-PI: [−0.15 - 0.17]), suggesting that more developed (in terms of what the HDI depicts) countries tend to exhibit less gender-based variability in creative thinking. Power distance was a positive moderator ($b = 0.002$, 95 %-CI [0.0007, 0.0030], $p < .05$; 95 %-PI: [−0.27 - 0.15]), indicating that in more hierarchical societies, male score dispersion may increase. In contrast, moderation analyses involving the GGGI were not statistically significant ($b = -0.465$, 95 %-CI [−1.06, 0.127], $p = .113$; 95 %-PI: [−0.27 - 0.07]).

Table 2
Results of meta-regressions including moderator variables.

Outcome	Predictor	Moderator	Est. (SE)	*p*	95 %-CI		95 %-PI		QM (df = 1)	τ^2
					LB	UB	*LB*	UB		
Crea	Gender	OECD	.025 (0.027)	.351	−0.029	.079	−0.488	−0.086	.859 ($p = .358$)	.009
		GGGI	.079 (0.360)	.829	−0.699	.857	−0.476	−0.059	.095 ($p = .759$)	.009
		HDI	.041 (0.235)	.862	−0.527	.444	−0.490	−0.053	.049 ($p = .825$)	.007
		PD	.001 (0.001)	.086	−0.000	.002	−0.458	−0.095	2.96 ($p = .093$)	.007
	SES	OECD	.024 (0.018)	.191	−0.012	.061	.179	.469	1.60 ($p = .211$)	.005
		GGGI	.264 (0.150)	.102	−0.060	.588	.210	.498	2.21 ($p = .142$)	.005
		HDI	.044 (0.001)	.461	−0.205	.293	.190	.483	.125 ($p = .725$)	.005
		PD	.000 (0.001)	.461	−0.001	.001	.200	.483	.559 ($p = .459$)	.004
	Gender*SES	OECD	.025 (0.009)	< 0.01	.006	.043	−0.096	.008	6.34 ($p < .05$)	.001
		GGGI	.036 (0.089)	.694	−0.154	.225	−0.073	.017	.152 ($p = .698$)	.000
		HDI	.293 (0.079)	< 0.01	.129	.456	−0.093	−0.005	24.1 ($p < .001$)	.000
		PD	−0.000 (0.000)	.270	−0.001	.000	−0.065	.028	2.65 ($p = .111$)	.000
Math	Gender	OECD	.118 (0.024)	< 0.001	.069	.116	−0.221	.177	20.0 ($p < .001$)	.009
		GGGI	.537 (0.275)	.073	−0.057	1.13	−0.172	.285	3.80 ($p = .056$)	.11
		HDI	.778 (0.199)	< 0.001	.368	1.19	−0.204	.193	21.5 ($p < .001$)	.009
		PD	−0.002 (0.001)	< 0.05	−0.003	−0.000	−0.093	.285	6.26 ($p < .05$)	.008
	SES	OECD	.060 (0.018)	< 0.05	.024	.095	.195	.477	10.5 ($p < .01$)	.005
		GGGI	.260 (0.177)	.165	−0.122	.641	.233	.528	2.16 ($p = .148$)	.005
		HDI	.224 (0.122)	.079	−0.028	.475	.207	.503	3.28 ($p = .076$)	.005
		PD	.000 (0.001)	.474	−0.000	.001	.251	.504	.550 ($p = .462$)	.003
	Gender*SES	OECD	.017 (0.012)	.137	−0.006	.040	−0.076	.048	2.45 ($p = .123$)	.001
		GGGI	−0.049 (0.116)	.682	−0.298	.200	−0.060	.047	.236 ($p = .629$)	.001
		HDI	.232 (0.114)	.052	−0.002	.466	−0.086	.047	9.55 ($p < .01$)	.001
		PD	−0.000 (0.000)	.455	−0.001	.001	−0.050	.063	1.49 ($p = .228$)	.001
Read	Gender	OECD	.075 (0.026)	< 0.01	.024	.126	−0.526	−0.125	7.83 ($p < .01$)	.010
		GGGI	.280 (0.296)	.362	−0.359	.919	−0.476	−0.071	1.29 ($p = .261$)	.009
		HDI	.489 (0.228)	< 0.05	.018	.960	−0.532	−0.105	7.37 ($p < .01$)	.010
		PD	−0.001 (0.001)	.213	−0.002	.000	−0.410	−0.079	1.40 ($p = .243$)	.006
	SES	OECD	.036 (0.017)	< 0.05	.003	.070	.190	.464	4.10 ($p < .05$)	.004

(*continued on next page*)

Table 2 (*continued*)

Outcome	Predictor	Moderator	Est. (SE)	*p*	95 %-CI		95 %-PI		QM (df = 1)	τ^2
					LB	UB	*LB*	UB		
		GGGI	.233 (0.170)	.194	−0.134	.601	.227	.227	2.02 (p = .161)	.004
		HDI	.065 (0.114)	.575	−0.170	.299	.210	.210	.324 (p = .571)	.004
		PD	.001 (0.000)	.217	−0.000	.001	.239	.239	1.35 (p = .252)	.003
	Gender*SES	OECD	.020 (0.012)	.053	−0.000	.041	−0.094	.024	3.73 (p = .058)	.001
		GGGI	−0.041 (0.112)	.723	−0.279	.198	−0.081	.034	.164 (p = .687)	.001
		HDI	.204 (0.096)	.096	.007	.402	−0.102	.026	8.22 (p < .01)	.001
		PD	−0.000 (0.000)	.592	−0.001	.001	−0.076	.040	.667 (p = .419)	.001

Notes. Crea = Creative Thinking, Math = Mathematics, Read = Reading, *SE* = Standard Error, 95 %-CI = 95 % Confidence Interval, 95 %-PI = 95 % Prediction Interval, LB = Lower Boundary, UB = Upper Boundary, *QM* = Moderator Test Statistic, τ^2 = Variance of True Effects. Gender coding: 0 = female, 1 = male.

Together, these findings suggest that while socio-demographic disparities in creative thinking remain stable across national contexts, even after introducing moderator variables, performance differences in mathematics and reading are more sensitive to structural and cultural characteristics of the respective countries and economies. Macro-level factors such as OECD membership and human development (as indicated through the HDI) appeared as the most important moderator variables, although their influence was not uniform across all predictors and domains. Instead, their effects varied depending on whether gender, SES, or their interaction was considered.

5. Discussion

Taken together, in this study we have used an innovative methodological approach (cf., Brunner et al., 2023) to analyze and integrate the data from 62 countries that have participated in the innovative domain *creative thinking* in the PISA 2022 study, as well as data from the two other crucial domains of PISA 2022: mathematics and reading. In sum, our results highlight the effects of gender and SES for the cognitive domains as well as for creative thinking as assessed in PISA 2022.

Regarding our first research question, we found that female students consistently outperformed male students in creative thinking across countries, with slightly greater score dispersion observed among male students; however, the magnitude of these differential effects varied considerably between the countries. A similar female advantage was found for reading, whereas male students showed a slight but significant advantage in mathematics. Regarding our second research question, we found that SES was a strong predictor of performance across all domains (including creative thinking), with higher SES correlating with better outcomes – this was true for two competing SES indices, that is, ESCS and HISEI (cf. Supplementary Materials). We observed that SES predicted creative thinking to a similar extent as it predicted performance in the other cognitive domains of mathematics and reading. The interaction terms (Gender*SES) were mostly negligible.

To contextualize the observed gender and SES effects in our study, it is helpful to consider typical benchmarks from the literature. Voyer and Voyer (2014) reported a mean gender difference in academic achievement of $d = 0.225$ (95 % CI [0.201, 0.249]), with the strongest female advantage found in reading-related domains. This aligns with our findings in both reading and creative thinking. For SES, meta-analyses have consistently shown moderate associations with academic achievement, typically ranging from $r = 0.22$ to $r = 0.29$ across subjects (Liu et al., 2022; Sirin, 2005), which closely matches the effect sizes observed in our analyses for creative thinking, mathematics, and reading.

Additionally, we examined whether country-level variables could account for cross-country variation in socio-demographic disparities in creative thinking and the other domains. Specifically, we tested four macro-level moderators—OECD membership, HDI, GGGI, and cultural power distance—across all models. Overall, the disparities in creative thinking associated with gender and SES were remarkably stable with regards to these different macro-level moderators and less susceptible to variation by these structural and cultural variables than those observed in mathematics and reading. In contrast to the stronger and more context-sensitive moderator effects found for the traditional academic domains (mathematics and reading), the consistency of effects in creative thinking suggests that social inequalities in this innovative domain may be existent due to more country-level structural characteristics.

5.1. What factors promote creative thinking?

Gender Differences and Their Implications. Based on the PISA 2022 assessment the short answer to the question which factors promote creative thinking might be: *Being Female and Being Well-Situated.* However, short answers often also fall short in grasping the

complexity of the issue at hand. Therefore, it is important that our findings are embedded into the previous literature and that awareness of potential limitations is raised.

First, the finding that female students outperform male students in creative thinking is consistent with previous meta-analytical findings in the literature (e.g., Abdulla Alabbasi et al., 2022; Taylor et al., 2024). However, the effect observed across diverse countries and economies participating in the PISA study was larger than those reported in the literature. This discrepancy may partly stem from differences in how creativity has been assessed: as a creative thinking skill versus as some form of achievement-related variable, or simply through self-reported behavior (Abraham, 2016; Baer & Kaufman, 2008). In light of the greater male variability hypothesis—which posits that males exhibit wider score distributions in cognitive traits, and in this case, creative thinking (cf., Taylor et al., 2024)—our results provide only limited support: while male students did show slightly higher variability in creative thinking, the effect was small and context-dependent. This suggests that although male dispersion may contribute a little to gendered patterns in creative thinking, it cannot fully account for the consistent female advantage observed at the mean level in the current global dataset. The wide prediction intervals—spanning near-zero and both negative and positive effects—highlight substantial residual heterogeneity across countries, suggesting that the strength and even direction of these associations may vary in future samples.

Remarkably in this context, the well-documented gender differences favoring females in creative thinking, which suggest greater creative potential for females, do not explain the disbalance in creative accomplishments or achievements, which tend to be more salient for males than for females (Abraham, 2016; Baer & Kaufman, 2008). Future research should therefore explore factors (such as rewards, as suggested by Abraham, 2016) that might contribute to gender differences, as well as how these differences in creative potential lead to creative behavior and, ultimately, creative achievements (Preckel et al., 2020; Benedek, 2024).

Another reason for the higher gender effects may be the specific age group of the PISA study. Adolescents, or emerging adults, are known to be undergoing substantial developmental dynamics regarding relevant traits such as task motivation which are known to be moderated by gender (Barbot et al., 2016). Finally, it would have been interesting to see whether gender differences varied by context domain (e.g., social versus scientific problem-solving) or whether these differences might be subject to effects based on expressional modalities. It remains an open question whether the observed gender differences would be more nuanced in verbal or figural expressional modalities. Although separate modality scores were not available at the time of the analysis, it seems promising to further investigate this assumption, because the observed gender differences in creative thinking were quite similar to those observed in reading. After all, one could speculate that creative thinking is affected due to differences in basic literacy levels.

Socio-Economic Status Effects and Their Interpretation. To better understand the observed cross-national variation in SES effects on creative thinking, it is helpful to reflect on how SES has been conceptualized and measured in educational research. Previously, SES has been assessed via different indicators or as a composite score of different indicators (e.g., Antonoplis, 2023; Avvisati, 2020). These indicators can include parental education, highest parental occupation (Highest International Socio-Economic Index of Occupational Status; *HISEI*), but also household goods such as books at home (Avvisati, 2020). While HISEI is a widely used and informative proxy for the economic standing of a family, more comprehensive indices such as the Economic, Social and Cultural Status (ESCS) index have been proposed to better capture the multidimensional nature of socio-economic background (Ehmke & Siegle, 2005). ESCS combines information on parental education, occupational status, and household possessions. Unlike single-indicator approaches, ESCS reflects the theoretical breadth of socio-economic background by integrating educational, occupational, and material resources. This multidimensional perspective has been shown to more reliably detect social disparities in educational outcomes and better account for variation in student achievement across systems, especially when the goal is to model cumulative socio-economic effects rather than isolate single pathways (Akukwe & Schroeders, 2016; Ehmke & Siegle, 2005). In line with this perspective, cultural capital—reflected in parental education and culturally relevant household items such as books and artworks—has been identified as a distinct component of SES with independent predictive value for educational success (Ehmke & Siegle, 2005; Tunmer, Chapman, & Prochnow, 2006). Beyond their material role, such cultural resources function as social signals that contribute to mechanisms of inclusion and exclusion in educational settings (Lamont & Lareau, 1988). Moreover, research has shown that cultural practices can serve as mediators between structural SES components and academic achievement, further supporting the need for multidimensional SES indicators (Baumert et al., 2003). Importantly, our use of the multidimensional ESCS index—as well as the consistent findings obtained with the HISEI—suggests that SES effects on creative thinking are robust across operationalizations, implying that the observed cross-country heterogeneity is unlikely to be driven solely by how SES is measured. This suggests that the impact of SES on educational outcomes varies across different countries or economies, maybe reflecting variations in how socio-economic advantages are translated into educational opportunities.

We observed SES as a strong predictor of performance across all cognitive domains. While previous meta-analytical evidence suggested that SES would predict performance in creative thinking, we expected it to be less influential as compared to its influence on mathematics and reading (Acar et al., 2022). This would have supported the notion that "necessity is the mother of invention“, but our findings did not reveal any differences in the prediction of creative thinking relative to the other cognitive domains. This can be considered unfortunate, because it means that educational goals, like gaining literacy in some cognitive domain, is across the board dependent on circumstances that one cannot choose for oneself (e.g., Li et al., 2023; Lindo, 2014). This stresses the importance of addressing socio-economic inequalities in education policy, particularly in ensuring that students from lower SES backgrounds receive adequate support to close performance gaps.

While SES showed a strong and consistent effect on creative thinking across countries, some degree of cross-national heterogeneity remained. However, none of the macro-level variables tested—such as OECD membership, HDI, GGGI, or power distance—significantly moderated this effect in the creative thinking domain, suggesting a globally persistent link between socio-economic advantage and performance. Likewise, gender-related mean differences in creative thinking remained largely unaffected by these contextual indicators. Interestingly, however, the variability of scores across gender—rather than the average gap—did vary

by macro-level features, including OECD status, HDI, and power distance. This indicates that while the direction and strength of gender effects are stable, national contexts still influence how nuanced the dispersion of students might be within each gender group.

Further possible explanatory variables might include broad contextual differences, such as variations between national educational systems or basic educational values. Additionally, more pragmatic issues like the extent to which the teaching of 21st century skills is implemented into a curriculum might be considered (Kain et al., 2024). Apart from such educational aspects, variables that rather tangent real-life circumstances in different countries, like access to extracurricular activities or cultural programs might be important (Yum, 2017). These ideas are also supported by recent ideas distinguishing between self-referenced and socially-referenced aspects of creativity (Dumas & Kaufman, 2024; Dumas et al., 2024). At age 15, lower SES students and males may primarily exhibit self-referenced creativity—ideas that are novel and useful to them but not necessarily to others—whereas the PISA assessment, scored by external raters, reflects socially-referenced creativity. This distinction suggests a developmental trajectory in which students may require increased domain knowledge and metacognitive skills to create ideas with broader social relevance, an area that may benefit from educational interventions targeting collaborative and knowledge-based creative activities.

Additionally, it would be valuable if future research would examine the effect of SES on other non-ability creativity measures included in the PISA 2022 study, such as individual beliefs about creativity and environmental factors like creative activities in schools. SES may play a different role in these areas as compared to performance measures. Although this was not the scope of the current work, we believe that this would reveal additional findings of educational interest. Interestingly, the minimal impact of OECD membership and related moderator variables on the relationship between gender and creative thinking performance suggests that factors related to economic development typically associated with OECD countries (OECD, 2024c), do not significantly influence gender differences in this domain. This conclusion was further strengthened using additional moderator variables on the macro-level. This calls for further investigation into other moderating variables, such as urban versus rural locations, parental involvement, peer influence, mental health and well-being, or educational quality that might explain these differences.

5.2. *Specific limitations of the assessment of creative thinking in PISA 2022*

The PISA 2022 study has been the first international effort to measure creative thinking in a large-scale setting across the globe. However, there are several points that should be discussed critically when it comes to the assessment of creative thinking in such an effort (for a recent call regarding opportunities, challenges, and cautions, cf. Barbot & Kaufman, 2025).

The theoretical structure of creative thinking comprised of three facets (generate diverse, generate creative and evaluate/improve ideas) as well as four contexts (written and visual expression, social and scientific problem solving) (OECD, 2023a); however, due to psychometrical considerations regarding scaling this theoretical structure was diminished to one total score reflecting creative thinking as an overarching construct. This modeling decision has not been empirically justified so far and it raises the question whether a total score obscures more nuanced insights. Furthermore, it can be criticized that based on the balanced incomplete block design as it is applied in PISA a number of students have only seen a few or even no creative thinking questions, but as typically carried out in PISA they have still received a plausible value. Future studies—as soon as scaling variants allow for it—should, therefore, focus on modeling creative thinking facets and contexts and should provide further evidence for the application of one value as well as for the effects that can be possible attributed to the balanced incomplete block design and the application of the plausible values.

Furthermore, it should be acknowledged that a highly time-extensive human coding of responses was realized, which, although considered a gold standard for DT scoring (Reiter-Palmon et al., 2019), differed notably from common procedures (Saretzki et al., 2024). Future studies should address this shortcoming and compare more frequently used scoring procedures, or even novel computerized scoring approaches based on large-language models (Goecke et al., 2024; Patterson et al., 2023), to the one applied here. Investigating scoring decisions might be helpful in preventing culture or country-specific coders' biases. In addition to that, while the PISA assessment aims for cross-cultural comparability, the standardized scoring of verbal tasks may still carry cultural-linguistic biases. As Barbot and Kaufman (2025) note, test content and scoring procedures—though technically uniform—may inadvertently reflect Western norms of expression, potentially disadvantaging responses that reflect other communicative styles. In this regard, figural tasks such as visual expression may be less culturally constrained and warrant separate consideration in cross-national comparisons.

Although the advisory group made extensive efforts to design the creative thinking assessments, it's important to recognize the substantial overlap between creative thinking and other cognitive domains (OECD, 2024a). This overlap may stem from basic competencies, such as reading, that are necessary for success across all PISA domains. However, from a multi-trait-multi-method perspective, the correlations between creative thinking and other cognitive domains are notably lower than the correlations among those other domains. This raises an interesting question about whether different facets or contexts of creative thinking correlate differently with the mathematics, reading, and science domains. Additionally, shared competencies like reading could influence relationships between creative thinking and factors such as SES or gender. An assessment of creative thinking that relies less on reading may come to different conclusions.

Lastly, given the overlap between creative thinking and the PISA core domains, it would be interesting to investigate outcome variables (e.g., creative achievements; Benedek, 2024) that might be solely/or to a higher degree explained by creative thinking than by the core domains. Such outcome variables can range from actual creative achievements (Benedek, 2024) to the choice of certain (potentially creative) occupations (e.g., based on O*Net; Cifuentes et al., 2010). Given the current results, it seems at least questionable to consider creative thinking measures somewhat less biased than measures of other cognitive domains, such as mathematics and reading (Kaufman, 2010) – although implementing such measures might certainly help to get a more holistic picture of a student taking those tests.

6. Conclusion

In summary, our analysis of data from 62 countries, using an integrative meta-analytical approach to individual participant data, reveals that both gender and SES significantly influence performance in creative thinking. Our findings suggest that female students consistently outperform male students in creative thinking, while SES remains a strong predictor of performance across all studied domains, including creative thinking, mathematics, and reading. Given these results, policymakers should prioritize addressing the substantial impact of SES on educational outcomes and the persistent gender disparities. Interventions aimed at reducing SES-related inequalities, such as providing targeted financial support, enhancing early childhood education, and ensuring access to quality educational resources, may be crucial not only for improving core academic domains but also for fostering creative thinking. These findings emphasize the importance of creating an educational environment that supports all students, regardless of gender or socio-economic background, in developing skills arguably necessary for success in the 21st century.

CRediT authorship contribution statement

B. Goecke: Writing – review & editing, Visualization, Validation, Supervision, Methodology, Investigation, Conceptualization. **M. Benedek:** Writing – review & editing, Validation, Investigation, Conceptualization. **J.G. Diedrich:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Investigation. **B. Forthmann:** Writing – original draft, Visualization, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **S. Patzl:** Writing – review & editing, Validation, Supervision, Software, Resources, Project administration, Methodology, Data curation, Conceptualization. **S. Weiss:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

Declaration of Conflict of Interest given their respective roles as Associate Editor and Co-Editor in Chief, Selina Weiss and Boris Forthmann had no involvement in the peer review of this article and had no access to information regarding its peer review. Full responsibility for the editorial process of this article was assigned to Co-Editor in Chief, Emmanuel Manalo. 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.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tsc.2025.101963.

Data availability

The data is openly available from the OECD. We make all analysis scripts openly available via the open science framework.

References

Abdulla Alabbasi, A. M., Thompson, T. L., Runco, M. A., Alansari, L. A., & Ayoub, A. E. A. (2022). Gender differences in creative potential: A meta-analysis of mean differences and variability. *Psychology of Aesthetics, Creativity, and the Arts*. https://doi.org/10.1037/aca0000506

Abraham, A. (2016). Gender and creativity: An overview of psychological and neuroscientific literature. *Brain Imaging and Behavior, 10*(2), 609–618. https://doi.org/10.1007/s11682-015-9410-8

Acar, S., Tadik, H., Uysal, R., Meyers, D., & Inetas, B. (2022). Socio-economic status and creativity: A meta-analysis. *Journal of Creative Behavior, 57*(1), 138–172. https://doi.org/10.1002/jocb.568

Akukwe, B., & Schroeders, U. (2016). Socio-economic, cultural, social, and cognitive aspects of family background and the biology competency of ninth-graders in Germany. *Learning and Individual Differences, 45*, 185–192. https://doi.org/10.1016/j.lindif.2015.12.009

Antonoplis, S. (2023). Studying socioeconomic status: Conceptual problems and an alternative path forward. *Perspectives on Psychological Science, 18*(2), 275–292.

Avvisati, F. (2020). The measure of socio-economic status in PISA: A review and some suggested improvements. *Large-Scale Assessments in Education, 8*(1), 8. https://doi.org/10.1186/s40536-020-00086-x

Baer, J., & Kaufman, J. C. (2008). Gender differences in creativity. *The Journal of Creative Behavior, 42*(2), 75–105. https://doi.org/10.1002/j.2162-6057.2008.tb01289.x

Barbot, B., & Kaufman, J. C. (2025). PISA 2022 creative thinking assessment: Opportunities, challenges, and cautions. *The Journal of Creative Behavior*. https://osf.io/78h6v.

Barbot, B., Lubart, T., & Besancon, M. (2016). Peaks, slumps, and bumps": Individual differences in the development of creativity in children and adolescents. *New Directions for Child and Adolescent Development, 151*, 33–45. https://doi.org/10.1002/cad.20152

Barra, C., & Boccia, M. (2022). What matters in educational performance? Evidence from OECD and non-OECD countries. *Quality & Quantity, 56*(6), 4335–4394. https://doi.org/10.1007/s11135-022-01322-y

Barth, P., & Stadtmann, G. (2024). Creativity in the West and the East: A meta-analysis of cross-cultural differences. *Creativity Research Journal*, 1–47. https://doi.org/10.1080/10400419.2024.2369442

Baumert, J., Watermann, R., & Schümer, G. (2003). Disparitäten der bildungsbeteiligung und des kompetenzerwerbs: Ein institutionelles und individuelles mediationsmodell. *Zeitschrift für Erziehungswissenschaft, 6*(1), 46–71. https://doi.org/10.1007/s11618-003-0004-7

Benedek, M. (2024). On the relationship between creative potential and creative achievement: Challenges and future directions. *Learning and Individual Differences, 110*, Article 102424. https://doi.org/10.1016/j.lindif.2024.102424
Benedek, M., Christensen, A. P., Fink, A., & Beaty, R. E. (2019). Creativity assessment in neuroscience research. *Psychology of Aesthetics, Creativity, and the Arts, 13*(2), 218–226. https://doi.org/10.1037/aca0000215
Bicer, A., Chamberlin, S., & Perihan, C. (2021). A meta-analysis of the relationship between mathematics achievement and creativity. *The Journal of Creative Behavior, 55*(3), 569–590. https://doi.org/10.1002/jocb.474
Blums, A., Belsky, J., Grimm, K., & Chen, Z. (2017). Building links between early socioeconomic status, cognitive ability, and math and science achievement. *Journal of Cognition and Development, 18*(1), 16–40. https://doi.org/10.1080/15248372.2016.1228652
Broer, M., Bai, Y., & Fonesca, F. (2019). A review of the literature on socioeconomic status and educational achievement. In M. Broer, Y. Bai, & F. Fonesca (Eds.), *Socioeconomic ineuality and educational outcomes* (pp. 7–17). Springer.
Brunner, M., Keller, L., Stallasch, S. E., Kretschmann, J., Hasl, A., Preckel, F., Lüdtke, O., & Hedges, L. V. (2023). Meta-analyzing individual participant data from studies with complex survey designs: A tutorial on using the two-stage approach for data from educational large-scale assessments. *Research Synthesis Methods*, 5–35. https://doi.org/10.1002/jrsm.1584
Charles, M., & Bradley, K. (2009). Indulging our gendered selves? Sex segregation by field of study in 44 countries. *American Journal of Sociology, 114*(4), 924–976. https://doi.org/10.1086/595942
Cheryan, S., Lombard, E. J., Hailu, F., Pham, L. N. H., & Weltzien, K. (2024). Global patterns of gender disparities in STEM and explanations for their persistence. *Nature Reviews Psychology, 4*(1), 6–19. https://doi.org/10.1038/s44159-024-00380-3
Cifuentes, M., Boyer, J., Lombardi, D. A., & Punnett, L. (2010). Use of O*NET as a job exposure matrix: A literature review. *American Journal of Industrial Medicine, 53*(9), 898–914. https://doi.org/10.1002/ajim.20846
Corazza, G. E., & de Saint-Laurent, C. (2020). Regional creativity: Research publications by region. *Encyclopedia of creativity* (3rd edition, p. 423). Elsevier, Academic Press. –238.
Cropley, A. J. (1967). *Creativity*. Longmans.
Delgadova, E. (2015). Reading literacy as one of the most significant academic competencies for the university students. *Procedia - Social and Behavioral Sciences, 178*, 48–53. https://doi.org/10.1016/j.sbspro.2015.03.145
Diedrich, J., Benedek, M., Jauk, E., & Neubauer, A. C. (2015). Are creative ideas novel and useful? *Psychology of Aesthetics, Creativity, and the Arts, 9*(1), 35–40. https://doi.org/10.1037/a0038688
Diedrich, J., & Lewalter, D. (2023). PISA 2022: Die Grundlagen. In D. Lewalter, J. Diedrich, F. Goldhammer, O. Köller, & K. Reiss (Eds.), *PISA 2022. Analyse der bildungsergebnisse in deutschland* (pp. 13–25). Waxmann Verlag GmbH.
Dumas, D., Forthmann, B., & Alexander, P. (2024). Using a model of domain learning to understand the development of creativity. *Educational Psychologist, 59*(3), 143–158. https://doi.org/10.1080/00461520.2023.2291577
Dumas, D., & Kaufman, J. C. (2024). Evaluation is creation: Self and social judgments of creativity across the four-C model. *Educational Psychology Review, 36*(4), 107. https://doi.org/10.1007/s10648-024-09947-1
Ehmke, T., & Siegle, T. (2005). ISEI, ISCED, HOMEPOS, ESCS. *Zeitschrift für Erziehungswissenschaft, 8*(4), 521–539. https://doi.org/10.1007/s11618-005-0157-7
Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136*(1), 103–127. https://doi.org/10.1037/a0018053
Enders, C. K., Du, H., & Keller, B. T. (2020). A model-based imputation procedure for multilevel regression models with random coefficients, interaction effects, and nonlinear terms. *Psychological Methods, 25*(1), 88–112. https://doi.org/10.1037/met0000228
Eriksson, K., Björnstjerna, M., & Vartanova, I. (2020). The relation between gender egalitarian values and gender differences in academic achievement. *Frontiers in Psychology, 11*, 236. https://doi.org/10.3389/fpsyg.2020.00236
Forgeard, M. J. C., & Kaufman, J. C. (2016). Who cares about imagination, creativity, and innovation, and why? A review. *Psychology of Aesthetics, Creativity, and the Arts, 10*(3), 250–269. https://doi.org/10.1037/aca0000042
Förster, N., Forthmann, B., Back, M. D., & Souvignier, E. (2023). Effects of the COVID-19 pandemic on reading performance of second grade children in Germany. *Reading and Writing, 36*(2), 289–315. https://doi.org/10.1007/s11145-022-10379-y
Gläveanu, V. P. (2019). Measuring creativity across cultures: Epistemological and methodological considerations. *Psychology of Aesthetics, Creativity, and the Arts, 13*(2), 227–232. https://doi.org/10.1037/aca0000216
Goecke, B., DiStefano, P. V., Aschauer, W., Haim, K., Beaty, R., & Forthmann, B. (2024). Automated scoring of scientific creativity in German. *The Journal of Creative Behavior*, jocb.658. https://doi.org/10.1002/jocb.658
Graciano, P., Lermen, F. H., Reichert, F. M., & Padula, A. D. (2023). The impact of risk-taking and creativity stimuli in education towards innovation: A systematic review and research agenda. *Thinking Skills and Creativity, 47*, Article 101220. https://doi.org/10.1016/j.tsc.2022.101220
Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology, 60*(1), 549–576. https://doi.org/10.1146/annurev.psych.58.110405.085530
Grund, S., Lüdtke, O., & Robitzsch, A. (2021). Multiple imputation of missing data in multilevel models with the R package mdmb: A flexible sequential modeling approach. *Behavior Research Methods, 53*(6), 2631–2649. https://doi.org/10.3758/s13428-020-01530-0
Guilford, J. P. (1966). Measurement and creativity. *Theory Into Practice, 5*(4), 185–189. https://doi.org/10.1080/00405846609542023
Guilford, J. P. (1967). *The nature of human intelligence*. McGraw-Hill.
Guilford, J. P. (1977). *Way beyond the IQ*. The Creative Education Foundation.
Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. *Science (New York, N.Y.), 320*(5880), 1164–1165. https://doi.org/10.1126/science.1154094
Guo, Y., Lin, S., Williams, Z. J., Grantham, T. C., Guo, J., Cole Clark, L. Q., & Zou, W. (2024). Creative potential and creative self-belief: Measurement invariance in cross-cultural contexts. *The Journal of Creative Behavior, 58*(2), 209–226. https://doi.org/10.1002/jocb.643
Guo, Y., Lin, S., Williams, Z. J., Zeng, Y., & Cole Clark, L. Q. (2023). Evaluative skill in the creative process: A cross-cultural study. *Thinking Skills and Creativity, 47*, Article 101240. https://doi.org/10.1016/j.tsc.2023.101240
Halpern, D. F. (2011). *Sex differences in cognitive abilities* (4th ed.). Psychology Press.
Harrington, D. M. (1975). Effects of explicit instructions to "be creative" on the psychological meaning of divergent thinking test scores [1]. *Journal of Personality, 43*(3), 434–454. https://doi.org/10.1111/j.1467-6494.1975.tb00715.x
He, J., & Van De Vijver, F. (2012). Bias and equivalence in cross-cultural research. *Online Readings in Psychology and Culture, 2*(2). https://doi.org/10.9707/2307-0919.1111
Hedges, L. V., & Friedman, L. (1993). Gender differences in variability in intellectual abilities: A reanalysis of Feingold's results. *Review of Educational Research, 63*(1), 94–105.
Heine, J.-H., Heinle, L., Hahnel, C., Lewalter, D., Becker-Mrotzek, M., & Lewalter, D. (2023). Lesekompetenz in PISA 2022: Ergebnisse, veränderungen und perspektiven. In J. Diedrich, F. Goldhammer, O. Köller, & K. Reiss (Eds.), *PISA 2022. Analyse der bildungsergebnisse in deutschland* (pp. 139–162). Waxmann Verlag GmbH.
Hernández-Torrano, D., & Ibrayeva, L. (2020). Creativity and education: A bibliometric mapping of the research literature (1975–2019). In *Thinking skills and creativity, 35*, Article 100625. https://doi.org/10.1016/j.tsc.2019.100625
Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*(11), 1539–1558. https://doi.org/10.1002/sim.1186
Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ (Clinical research ed.), 327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the mind: Intercultural cooperation and its importance for survival (Revised and expanded third edition)*. McGraw-Hill.
Huber, S. J., & Paule-Paludkiewicz, H. (2024). Gender norms and the gender gap in higher education. *Labour Economics, 87*, Article 102491. https://doi.org/10.1016/j.labeco.2023.102491
Ivcevic, Z., & Grandinetti, M. (2024). Artificial intelligence as a tool for creativity. *Journal of Creativity, 34*(2), Article 100079. https://doi.org/10.1016/j.yjoc.2024.100079
Jaquish, G. A., & Ripple, R. E. (1985). A life-span developmental cross-cultural study of divergent thinking abilities. *The International Journal of Aging and Human Development, 20*(1), 1–11.
Javidan, M., House, R. J., Dorfman, P. W., Hanges, P. J., & Sully De Luque, M. (2006). Conceptualizing and measuring cultures and their consequences: A comparative review of GLOBE's and Hofstede's approaches. *Journal of International Business Studies, 37*(6), 897–914. https://doi.org/10.1057/palgrave.jibs.8400234
Jellen, H. G., & Urban, K. K. (1989). Assessing creative potential world-wide: The first cross-cultural application of the test for creative thinking—drawing production (TCT-DP). *Gifted Education International, 6*(2), 78–86. https://doi.org/10.1177/026142948900600204
Kain, C., Koschmieder, C., Matischek-Jauk, M., & Bergner, S. (2024). Mapping the landscape: A scoping review of 21st century skills literature in secondary education. *Teaching and Teacher Education, 151*, Article 104739. https://doi.org/10.1016/j.tate.2024.104739
Karwowski, M. (2016). Culture and psychometric studies of creativity. *The Palgrave handbook of creativity and culture research* (pp. 159–186).
Kaufman, J. C. (2010). Using creativity to reduce ethnic bias in college admissions. *Review of General Psychology, 14*(3), 189–203. https://doi.org/10.1037/a0020133
Kornilov, S. A., Kornilova, T. V., & Grigorenko, E. L. (2016). The cross-cultural invariance of creative cognition: A case study of creative writing in U.S. and Russian college students: The cross-cultural invariance of creative cognition. *New Directions for Child and Adolescent Development, 2016*(151), 47–59. https://doi.org/10.1002/cad.20149
Lamont, M., & Lareau, A. (1988). Cultural capital: Allusions, gaps and glissandos in recent theoretical developments. *Sociological Theory, 6*(2), 153. https://doi.org/10.2307/202113
Lewalter, D., Diedrich, J., Goldhammer, F., Köller, O., & Reiss, K. (Eds.). (2023). *PISA 2022: Analyse der bildungsergebnisse in deutschland*. Waxmann Verlag GmbH. https://doi.org/10.31244/9783830998488.
Li, J., Peng, P., Ma, X., Ding, N., & Zhao, J. (2023). How does family socioeconomic status influence children's reading ability? Evidence from meta-analytic structural equation modeling. *Educational Psychology Review, 35*(4), 119. https://doi.org/10.1007/s10648-023-09834-1
Lindo, E. J. (2014). Family background as a predictor of reading comprehension performance: An examination of the contributions of human, financial, and social capital. *Learning and Individual Differences, 32*, 287–293. https://doi.org/10.1016/j.lindif.2014.03.021
Liu, J., Peng, P., Zhao, B., & Luo, L. (2022). Socioeconomic status and academic achievement in primary and secondary education: A meta-analytic review. *Educational Psychology Review, 34*(4), 2867–2896. https://doi.org/10.1007/s10648-022-09689-y
Lubart, T., Pacteau, C., Jacquet, A.-Y., & Caroff, X. (2010). Children's creative potential: An empirical study of measurement issues. *Learning and Individual Differences, 20*(4), 388–392. https://doi.org/10.1016/j.lindif.2010.02.006
Lüdtke, O., Robitzsch, A., & West, S. G. (2020). Regression models involving nonlinear effects with missing data: A sequential modeling approach using Bayesian estimation. *Psychological Methods, 25*(2), 157–181. https://doi.org/10.1037/met0000233
Mcgranahan, D., & Wojan, T. (2007). Recasting the creative class to examine growth processes in rural and urban counties. *Regional Studies, 41*(2), 197–216. https://doi.org/10.1080/00343400600928285
Müller, K., & Ehmke, T. (2013). Soziale herkunft als bedingung derKompetenzentwicklung. In M. Prenzel, C. Sälzer, E. Klieme, & O. Köller (Hrsg.) (Eds.), *PISA 2012. Fortschritte und Herausforderungen in Deutschland* (pp. 245–275). Münster: Waxmann.
Nguyen, T., Watts, T. W., Duncan, G. J., Clements, D. H., Sarama, J. S., Wolfe, C., & Spitler, M. E. (2016). Which preschool mathematics competencies are most predictive of fifth grade achievement? *Early Childhood Research Quarterly, 36*, 550–560. https://doi.org/10.1016/j.ecresq.2016.02.003
Nieminen, P. (2022). Application of standardized regression coefficient in meta-analysis. *BioMedInformatics, 2*(3), 434–458. https://doi.org/10.3390/biomedinformatics2030028
Niu, W., & Kaufman, J. C. (2013). Creativity of chinese and american cultures: A synthetic analysis. *The Journal of Creative Behavior, 47*(1), 77–87. https://doi.org/10.1002/jocb.25
Niu, W., & Sternberg, R. J. (2001). Cultural influences on artistic creativity and its evaluation. *International Journal of Psychology, 36*(4), 225–241. https://doi.org/10.1080/00207590143000036
OECD. (2012). Closing the gender gap: Act now. *OECD*. https://doi.org/10.1787/9789264179370-en
OECD. (2019). PISA 2018 reading framework. *PISA 2018 assessment and analytical framework*. OECD Publishing. https://doi.org/10.1787/5c07e4f1-en
OECD. (2023a). PISA 2022 creative thinking framework. *PISA 2022 assessment and alaytical framework*. OECD Publishing. https://doi.org/10.1787/471ae22e-en
OECD. (2023b). PISA 2022 mathematics framework. *PISA 2022 assessment and analytical framework*. OECD Publishing. https://doi.org/10.1787/7ea9ee19-en
OECD. (2024a). *PISA 2022 results creative minds, creative schools volume iii*. OECD Publishing. https://doi.org/10.1787/765ee8c2-en
OECD. (2024b). *PISA 2022 technical report*. OECD Publishing. https://www.oecd.org/pisa/data/pisa2022technicalreport.
OECD. (2024c,. September). *Convention on the organisation for economic co-operation and development*.
Patterson, J. D., Barbot, B., Lloyd-Cox, J., & Beaty, R. E. (2023). AuDrA: An automated drawing assessment platform for evaluating creativity. *Behavior Research Methods, 56*(4), 3619–3636. https://doi.org/10.3758/s13428-023-02258-3
Pichot, N., Forthmann, B., Bonetto, E., Arciszewski, T., Bonnardel, N., Jaubert, S., & Pavani, J. B. (2024). Aiming at creativity and ending up with a range from low-hanging fruits to foolishness: A reflective model of creativity. *The Journal of Creative Behavior*, jocb.667. https://doi.org/10.1002/jocb.667
Plucker, J. A., Beghetto, R. A., & Dow, G. T. (2004). Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educational Psychologist, 39*(2), 83–96. https://doi.org/10.1207/s15326985ep3902_1
Preckel, F., Golle, J., Grabner, R., Jarvin, L., Kozbelt, A., Müllensiefen, D., et al. (2020). Talent development in achievement domains: A framework for within- and cross-domain research. *Perspectives on Psychological Science, 15*(3), 691–722.
Puccio, G. J. (2017). From the dawn of humanity to the 21st century: Creativity as an enduring survival skill. *The Journal of Creative Behavior, 51*(4), 330–334. https://doi.org/10.1002/jocb.203
Pustejovsky, J. E., & Tipton, E. (2022). Meta-analysis with robust variance estimation: Expanding the range of working models. *Prevention Science, 23*(3), 425–438. https://doi.org/10.1007/s11121-021-01246-3
Rafner, J., Beaty, R. E., Kaufman, J. C., Lubart, T., & Sherson, J. (2023). Creativity in the age of generative AI. *Nature Human Behaviour, 7*(11), 1836–1838. https://doi.org/10.1038/s41562-023-01751-1
Reilly, D. (2012). Gender, culture, and sex-typed cognitive abilities. *PLoS ONE, 7*(7), Article e39904. https://doi.org/10.1371/journal.pone.0039904
Reiter-Palmon, R., Forthmann, B., & Barbot, B. (2019). Scoring divergent thinking tests: A review and systematic framework. *Psychology of Aesthetics, Creativity, and the Arts, 13*(2), 144–152. https://doi.org/10.1037/aca0000227
Ritchie, S. J., Luciano, M., Hansell, N. K., Wright, M. J., & Bates, T. C. (2013). The relationship of reading ability to creativity: Positive, not negative associations. *Learning and Individual Differences, 26*, 171–176. https://doi.org/10.1016/j.lindif.2013.02.009
Robitzsch, A., & Lüdtke, O. (2024). *mdmb: Model based treatment of missing data* (p. 1.9–22) [Computer software]. https://CRAN.R-project.org/package=mdmb.
Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*(3), 581–592.
Runco, M. A. (1991). The evaluative, valuative, and divergent thinking of children*. *The Journal of Creative Behavior, 25*(4), 311–319. https://doi.org/10.1002/j.2162-6057.1991.tb01143.x
Runco, M. A., & Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creativity Research Journal, 24*(1), 66–75. https://doi.org/10.1080/10400419.2012.652929
Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal, 24*(1), 92–96. https://doi.org/10.1080/10400419.2012.650092

Rutkowski, L., Gonzalez, E., Joncas, M., & Von Davier, M. (2010). International large-scale assessment data: Issues in secondary analysis and reporting. *Educational Researcher, 39*(2), 142–151. https://doi.org/10.3102/0013189X10363170
Rutkowski, L., Gonzalez, E., von Davier, M., & Zhou, Y. (2013). Assessment design for international large-scale assessments. *Handbook of international large-scale assessment*. Chapman and Hall/CDC.
Said-Metwaly, S., Fernández-Castilla, B., Kyndt, E., & Van Den Noortgate, W. (2020). Testing conditions and creative performance: Meta-analyses of the impact of time limits and instructions. *Psychology of Aesthetics, Creativity, and the Arts, 14*(1), 15–38. https://doi.org/10.1037/aca0000244
Saretzki, J., Forthamnn, B., & Benedek, M. (2024). A systematic quantitative review of divergent thinking assessments. *Psychology of aesthetics, creativity, and the arts, advanced online publication.* https://doi.org/10.1037/aca0000691
Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417–453. https://doi.org/10.3102/00346543075003417
Spinath, B., Eckert, C., & Steinmayr, R. (2014). Gender differences in school success: What are the roles of students' intelligence, personality and motivation? *Educational Research, 56*(2), 230–243. https://doi.org/10.1080/00131881.2014.898917
Stanat, P., Schipolowski, S., Schneider, R., Sachse, K., Weinrich, S., & Henschel, S. (2022). *QB-bildungstrend 2021. Kompetenzen in den fächern deutsch und Mathema- tik am ende der 4. jahrgangsstufe im dritten ländervergleich.* Waxmann.
Stoet, G., & Geary, D. C. (2013). Sex differences in mathematics and reading achievement are inversely related: Within- and across-nation assessment of 10 years of PISA data. *PLoS ONE, 8*(3), Article e57988. https://doi.org/10.1371/journal.pone.0057988
Taylor, C. L., Said-Metwaly, S., Camarda, A., & Barbot, B. (2024). Gender differences and variability in creative ability: A systematic review and meta-analysis of the greater male variability hypothesis in creativity. *Journal of Personality and Social Psychology, 126*(6), 1161–1179. https://doi.org/10.1037/pspp0000484
The Global Gender Gap Report 2022. (2023). World economic forum.
Tunmer, W. E., Chapman, J. W., & Prochnow, J. E. (2006). Literate cultural capital article. *New Zealand Journal of Educational Studies, 41*(2), 183–204.
United Nations Development Programme. (2025). *Human development report 2025: A matter of choice: People and possibilities in the age of AI*. United Nations.
Viechtbauer, W. (2010). Conducting meta-analyses in *R* with the metafor package. *Journal of Statistical Software, 36*(3). https://doi.org/10.18637/jss.v036.i03
von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful. *IERI Monograph Series, 2*(1), 9–36.
Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140*(4), 1174–1204. https://doi.org/10.1037/a0036620
Wang, A. Y. (2012). Exploring the relationship of creative thinking to reading and writing. *Thinking Skills and Creativity, 7*(1), 38–47. https://doi.org/10.1016/j.tsc.2011.09.001
Weiss, S., Wilhelm, O., & Kyllonen, P. (2021). An improved taxonomy of creativity measures based on salient task attributes. *Psychology of Aesthetics, Creativity, and the Arts.* https://doi.org/10.1037/aca0000434. Advance online publication.
Wilson, R. C., Guilford, J. P., & Christensen, P. R. (1953). The measurement of individual differences in originality. *Psychological Bulletin, 50*(5), 362.
Yum, S. (2017). The spatial patterns of creativity in the U.S. metropolitan areas. *Business Creativity and the Creative Economy, 3*, 99–109. https://doi.org/10.18536/bcce.2017.10.3.1.09