Reading ability conflates SES creativity gaps

Guilherme Lichand[a], **Lêticia Lopes**[b], **and Sachin Allums**[c]

This manuscript was compiled on January 29, 2026

Who is more creative: high- or low-socioeconomic status (SES) individuals? This question is the focus of intense debates within both the scientific community and society at large, since creativity has been linked to innovation, productivity, and wealth generation. This question is, however, hard to answer; in particular, because creativity is typically assessed through standardized tests that build on reading and writing proficiency, which might conflate the relationship between SES and creativity. To overcome this challenge, we combine high-quality data on reading ability and experimental variation in reading requirements embedded in creativity assessments, in a series of studies with 6-12th graders in Brazilian schools. We first document that established creativity measures exhibit sizable SES gaps, but that these gaps become much smaller and no longer significant once we parse out students' reading ability (Study 1). Next, in two randomized control trials, we have students complete divergent thinking tasks while experimentally varying reading requirements: in the control group, students had to read question prompts, just like in standard assessments (e.g., PISA); in the treatment group, enumerators read the prompts on their behalf. In both conditions, enumerators wrote down students' answers. We find that high-SES students outperform low-SES ones *only* when they read the prompt, but not when performance does not depend on reading ability (Studies 2 and 3).

Creativity | SES gaps | Reading ability | Test unfairness

Significance Statement

We document that, while established creativity measures exhibit sizable socioeconomic status (SES) gaps, these gaps become much smaller and no longer significant once we parse out students' reading ability. Leveraging two randomized control trials, we find that high-SES students outperform low-SES ones in divergent thinking tasks *only* when assessments embed reading requirements.

1. Introduction

Are high-SES individuals more creative than low-SES ones? There are compelling reasons to think this might be the case. Numerous studies find positive correlations between SES and creative performance (1–3), a pattern that holds even across cultures (4). SES tends to positively correlate with executive functions, which in turn, could have downstream effects on creative potential (5). SES could further boost subjects' confidence in their creative abilities, or their creative self-efficacy, which could then positively impact their creative skills (6). Creativity is often linked to innovation, entrepreneurship, productivity, and ultimately, wealth creation; thus, the most creative individuals may be those who accumulate wealth. Moreover, wealth might provide cognitive freedom, freeing people from immediate survival concerns and allowing them to think more imaginatively and beyond conventional constraints (3, 6, 7). In contrast, there are plausible mechanisms that might lead the exact opposite to hold. More creative individuals may feel unfulfilled by the repetitive tasks typical in many educational and workplace settings – with potential detrimental effects to their productivity, learning, and employability, particularly within societies that undervalue creative skills. Moreover, limited resources can push individuals to think creatively out of necessity, developing resourceful ways to meet basic needs (3, 8). Determining which effect dominates (if any) is an empirical question, widely debated in both academic research and public discourse.

The ascribed relevance of creative thinking, especially in a context of increasing automation since the advent of generative AI (9), has motivated recent efforts to measure creativity globally, such as the extensive Program for International Student Assessment (PISA) framework, which assesses 15-year-olds in over 60 countries. The 2022 exam specifically evaluated students' abilities to generate diverse and original

Author affiliations: [a]Stanford Graduate School of Education, Stanford, CA, United States. Corresponding author: glichand@stanford.edu.; [b]São Paulo School of Economics (FGV-EESP), São Paulo, Brazil; [c]Stanford University, Stanford, CA, United States

ideas, as well as to assess and refine ideas across various contexts, using open-ended tasks focused on communication and problem-solving. The OECD report on PISA creativity scores documents a sizable socioeconomic status (SES) gap, both across and within countries. Concretely, the gap between the high-income OECD countries and Brazil (the setting of our study) was 10 points in the PISA scale – 83% of a standard deviation of OECD students' creative thinking performance. This difference is comparable to the SES gap within Brazil, where low-SES students scored 11 points lower than their high-SES counterparts – a nearly 50% difference in average scores. According to the report, a 1-unit increase in the SES index within Brazilian 15-year-olds assessed by PISA corresponds to a 4-point improvement in creative thinking performance (10).

At face value, these correlations support the theory that creativity is positively correlated with wealth. However, performing well on the PISA creativity assessment – a standard assessment, based on open-ended questions – *also requires* strong reading and writing skills. Since low-SES students often have lower proficiency in these areas (3), the observed SES gap in creativity scores may, at least in part, reflect underlying disparities in reading and writing ability.

We report the results of three studies that address this challenge, conducted with different samples of 6-12th graders in Brazilian schools. Study 1 first documents SES gaps in standard creativity measures and investigates the extent to which reading ability mediates these gaps. Study 2 then explores whether experimentally removing reading requirements from the creativity assessments used in Study 1 closes these SES gaps. Finally, Study 3 replicates that experimental manipulation in the context of items from the PISA 2022 Creative Thinking assessment.

This paper hopes to establish whether SES gaps in creativity measures can be attributed to reading requirements rather than to meaningful differences in creative potential. This work adds to an ongoing debate about SES gaps in creativity, with implications for the assessment of higher-order skills in general, and creativity in particular.

2. Results

Study 1: SES creativity gaps in standard creativity measures conflate reading ability. In line with evidence from international assessments like PISA and with previous research findings, Figure 1 documents a sizable and statistically significant SES creativity gap based on the summary measure of conventional creativity measures (the AUT and the DAT). Concretely, high-SES students score 0.154 s.d. higher than low-SES students (p-value of difference = 0.023).

Figure 2 documents that, however, this gap at least partly conflates differences in reading ability across low- and high-SES participants. Leveraging high-quality reading ability data from ROAR-SRE, we document that while the SES creativity gap among proficient students is of similar magnitude (0.150 s.d.), albeit imprecisely estimated (p-value of difference = 0.076), such gap is 1/3 lower among those who do not read at grade-5 level (0.098 s.d.) and no longer statistically significant (p-value of difference = 0.324). Supplementary Materials document that results hold for the different components of the summary measure.

It is also striking that, even among students who read proficiently, the SES creativity gap is 2/3 lower than the creativity gap across reading ability levels in the whole sample (0.470 s.d., p-value of difference < 0.001).

Together, the patterns we document are consistent with the claim that reading ability conflates SES creativity gaps in standard creativity measures.

Nonetheless, since reading ability is not randomly assigned – but, rather, correlated with several student characteristics, including household wealth –, it could be that our findings indicate that our reading ability measure is merely a better proxy for SES than the one we rely on (number of full bathrooms). For this reason, Studies 2 and 3 introduce experimental variation in reading requirements embedded in creativity assessments.

Study 2: High-SES students outperform low-SES ones in standard creativity measures only when they read the prompts. Figure 3 documents that high-SES students outperform low-SES ones in standard creativity measures (the AUT and the DAT) only due to reading requirements.

Among participants who had to read the question prompts, we estimate a SES creativity gap similar to the unconditional one documented in Study 1 (0.134 s.d., p-value of the difference = 0.023). Conversely, among those for whom enumerators read the prompts on their behalf, the SES gap is less than 1/3 (0.068 s.d.) and no longer statistically significant (p-value of the difference = 0.249). Supplementary Materials document that results hold for the different components of the summary measure.

The experimental findings provide causal evidence that SES gaps in conventional creativity measures indeed conflate differences in reading ability.

Next, Study 3 investigates whether the same holds for less conventional creativity measures included in international assessments like PISA.

Study 3: High-SES students outperform low-SES ones in PISA-like creativity measures only when they read the prompts. Figure 4 documents that, similar to Study 2, reading requirements mask differences between high- and low-SES students in PISA-like creativity measures (diverse and creative idea generation tasks).

Among participants who had to read the question prompts, we estimate virtually no SES creativity gap (a 0.004 s.d. difference in favor of high-SES students, p-value = 0.903). In turn, among those for whom enumerators read the prompts on their behalf, a sizeable SES gap emerges *in favor of low-SES students* (0.085 s.d., p-value = 0.011).

Supplementary Materials replicate the analyses for the different components of the summary measure. Consistent with the main text, SES differences in the share of students who receive full credit in the first item (based on the PISA's scoring guidelines) are reversed in the absence of reading requirements (p-value of difference-in-difference = 0.043).

Supplementary Materials further show that, when using GenAI instead of hand-coded scores, the patterns of SES gaps match precisely those documented in Study 2: high-SES students *only* outperform low-SES ones in the presence of reading requirements.

3. Discussion

Together, our findings provide compelling evidence that reading ability conflates creativity measures in standard assess-

ments – including PISA 2022 –, in particular when it comes to SES gaps. The OECD report acknowledges that "[a]fter accounting for students' mathematics and reading performance, differences in the performance of advantaged and disadvantaged students are much smaller in all countries/economies – and even become statistically non-significant in 14 [of them]" (11, p. 113). While it posits that "[S]ocio-economic disparities in creative thinking performance therefore rather reflect a range of economic and cultural factors, experiences and mechanisms known to affect student achievement overall" (11, p. 113), it does not mention test unfairness as a potential driver of SES differences. Instead, the report concludes that, since the correlation between reading and creativity scores across OECD countries is 0.66 – less than that between reading and other disciplines –, "the creative thinking assessment measures a different subset of skills with respect to those measured in the mathematics, reading and science assessments" (11, p. 83). This study challenges this conclusion; concretely, it suggests that SES differences in reading comprehension can account for the *totality* of SES differences in average creativity scores across income groups within countries, and across countries of different income levels.

More broadly, while previous research documents a positive correlation between SES gaps in literacy skills and those in other subjects, like math (12), it is hard to infer unfairness solely based on these associations. As (13) points out, "[a] major limitation of item bias statistics or indices is that measures of relative difficulty do not provide proof of unfairness. Only if an item is relatively more difficult for one group (statistically biased) *and* the source of this difficulty is irrelevant to the test construct is an item said to be unfair" (13, p. 234). Our paper innovates by experimentally varying whether test performance hinges on reading proficiency – a source of difficulty irrelevant to the divergent thinking constructs that the PISA aims to capture. Having enumerators write down responses for participants further allows us to isolate creativity from writing ability. All in all, we provide causal evidence that the control version of the test is *unfair*.

An active literature (e.g., 14, 15) discusses how creativity might reflect a range of underlying factors, from general intelligence to other latent variables. If both reading ability and creativity reflect cognition more broadly (construct redundancy), one could ask whether it is surprising at all that when we control for one (as in Study 1) we partially eliminate variation in the other. This is not, however, what we do in Studies 2 and 3, which experimentally isolate the role of reading requirements – not that of reading ability or its underlying factors – in expressing creative ability.

Our findings are naturally limited by our study design. Even if Studies 1 and 2 conducted the same tests, differences in student population, assessment type, and time limits constrain our ability to directly compare SES gaps across them. In particular, Study 1 was self-administered: participants had 2 minutes to complete the AUT and 4 minutes to complete the DAT (computer-based) as they had to write down their responses. In contrast, Studies 2 and 3 were mediated by surveyors: across experimental conditions, participants had only 30 seconds to complete the AUT and 45 seconds to complete the DAT (tablet-based) because enumerators wrote responses on their behalf. Despite these differences in administration and sampling (a non-random sample of Rio middle-schoolers in Study 1, and a representative sample of Brazilian K-12 students in Studies 2 and 3), we observe SES creativity gaps of very similar magnitudes in the AUT and DAT in both cases.

We carefully designed the three studies to ensure our findings are not country-specific. Studies 1 and 2 featured the AUT, the most widely used divergent thinking task in the creativity literature (16), and the DAT, a task with an objective scoring methodology, unlike many other divergent thinking tasks (16, 17). We chose these measures in particular because they have been traditionally used as evidence that high-SES subjects might be more creative than low-SES ones (3). The fact that we find similar patterns across the lab-in-the-field AUT and DAT tests in Studies 1 and 2 and the PISA items in Study 3 (which presumably align more closely to real-world creative problem solving) lends some confidence that our findings are relevant above and beyond the artificial setting of our experiments. While we use fluency as the primary measure of creativity on the AUT, Supplementary Materials show that our conclusions extend to other ways of scoring the AUT, such as originality, which might more closely measure true creative potential instead of merely conflating generating ideas with creative ability (18). The items used in Study 3 were extracted from PISA itself, and the SES creativity gaps in our control group correspond almost exactly to those documented in the OECD report. Brazil was the setting of this study thanks to the unique opportunity to conduct large-scale survey experiments in more than 200 schools throughout the country – a research-practice partnership that is hard to replicate anywhere else.

Despite these efforts, replicating these experiments in other settings and with additional creativity measures to ensure that these results generalize further remains a promising avenue for future work. Comparative education research directly contrasting the performance of high-income OECD countries and that of lower-income countries like Brazil, featuring the experimental manipulation of Studies 2 and 3, could provide convincing evidence to challenge the findings of the OECD report. It also remains unclear if these findings would extend to other aspects of verbal creativity (i.e., convergent thinking), other assessments for divergent thinking (i.e., form-first vs function-first tasks), or even figural creativity (completing abstract figures in a novel way; e.g., 19). As a leading example, while reading instructions aloud could help in figural creativity tasks, since participants might understand more clearly that they should not simply complete an obvious image, we cannot say for certain whether that would be the case. Future work could focus on using additional measures to examine whether reading ability also conflates SES creativity gaps in those cases.

Overall, our findings suggest that assessments of higher-order skills, such as creative thinking, should better account for limitations in foundational skills (often disproportionately concentrated among low-income students). Specifically, assessors might consider administering oral exams or even non-verbal measures of creativity or other higher-order skills in order to accurately capture underlying skills. More broadly, in line with factor-analytic frameworks (e.g., 14, 15, 20), our findings suggest that reading requirements embedded in conventional assessments might conflate a host of cognitive abilities, from executive functions to mathematical reasoning. Ultimately,

unfairness in test design not only constrains the extent to which the assessment accurately reflects levels and differences in the skills or psychological constructs it intends to capture, but also, might end up reinforcing deficit narratives that mischaracterize the potential of disadvantaged students.

Methods

Study 1 was approved by the Stanford University Institutional Review Board (protocol #77799) on May 29, 2025. Studies 2 and 3 were approved by the Brazilian Institutional Review Board (CONEP; protocol #7.168.931) on October 18, 2024, and by the Stanford University Institutional Review Board (protocol #77912) on January 17, 2025. Waiver of parental consent obtained from all school principals, as the studies involved no more than minimal risk to study participants, whom still had to assent and could drop out of the study at any point and without penalty. Supplementary Materials contain additional descriptive statistics information on the study samples and robustness checks for our findings.

Study 1: Documenting SES creativity gaps and the extent to which reading ability mediates them. In May-Jul/2025, the study enrolled 523 students across 8 middle schools (grades 6-9) in the municipality of Rio de Janeiro, Brazil. Recruited students participated in computer-based reading and creativity assessments, covering conventional divergent thinking tasks. Divergent thinking is a core component of creative thinking: the ability to generate a variety of creative possibilities and novel associations (21, 22).

The analyses proxy for SES by comparing students living in households with at most one full bathroom (*low-SES*) to those with two or more (*high-SES*). This is considered a reliable proxy for SES in contexts where direct measures of income or wealth are unavailable or impractical (23). Since children were the ones filling out the survey with no caregivers present, we decided to use full bathrooms as a proxy for SES. Based on this proxy, 366 study participants were low-SES and 157 were high-SES – a distribution very similar to the Brazilian national figures, based on the 2022 Census (24).

We measure students' reading ability through the Rapid Online Assessment of Reading (ROAR; 25–29), a digital assessment that measures reading fluency and comprehension. Concretely, we assessed participants' reading fluency and comprehension using the ROAR Sentence Reading Efficiency (ROAR-SRE) task, previously validated for Brazil (29). In the task, participants rate sentences as true or false as quickly and accurately as they can, within 180 seconds. Sentences were carefully validated for Brazilian Portuguese and screened for contextual suitability, requiring minimal background knowledge and using simple vocabulary and syntactic structures (29). Sentences are presented in random order. We score the task as the number of correct answers minus the number of incorrect answers over that time span, in line with (29).

Based on (29), which associates ROAR-SRE scores to proficiency labels in Brazil, 43.02% (N = 225) of 6-9th graders in our sample do not read at grade-5 level (43.44% among low-SES students, and 42.04% among high-SES ones).

We conducted five trials of the Alternative Uses Task (AUT) and one trial of the Divergent Association Task (DAT) with all study participants in the context of a companion paper (30), which randomly assigned a generative AI tool to assist students with the creativity tasks. In this study, we restrict attention to the AUT unassisted trials (the first and the last ones), and parse out treatment indicators across all analyses.

In the AUT, participants have 2 minutes to come up with as many creative and unusual uses for an everyday object. Participants completed the AUT for the five following items: tire, pants, shoe, table, and bottle. These objects were chosen because their uses have the highest correlation of judgements between either an automated or human evaluator for originality (31). The order of the objects was randomly assigned across participants.

The AUT is typically scored on different dimensions: *fluency* (the number of distinct uses), *flexibility* (the number of categories to which alternative uses correspond), and *originality* (responses' statistical infrequency) – each deemed as a key indicator of creative potential (21, 22, 32, 33). In the main text, we restrict attention to fluency, the most common and transparent dimension used in the literature; Supplementary Materials replicate all analyses using originality instead of fluency.

In the DAT, participants have 4 minutes to come up with 10 nouns as semantically different as possible from one another (17). For example, 'dog' and 'cat' are very close in meaning, and thus, including both of them would result in a low DAT score. By contrast, 'dog' and 'guitar' would score much higher since they have very different meanings and uses. In the DAT, words must be single words, non-proper nouns, and not be considered technical vocabulary. Only the first seven valid words are scored, following the scoring algorithm in (17). If a participant did not provide seven valid words, they received a score of zero.

Participants could refuse answering one or more trials; response rates for each trial were statistically identical across SES (92.1% for low-SES and 91.7% for high-SES, p-value of difference = 0.848).

We compute p-values from Wald tests of differences in average test scores between high- and low-SES participants, at first, unconditional on reading ability, and next, conditioning on whether participants could read at grade-5 level (*high literacy*) or not (*low literacy*). To deal with family-wise error rates in hypotheses testing from multiple comparisons, we focus on a divergent thinking *summary measure*, stacking the dataset with participants' test scores in each trial (2 observations for AUT and 1 for DAT) and clustering the analyses at the student level. We also control for age, gender, and treatment assignment in the companion paper. Supplementary Materials showcase results without controls, and for each component of the summary measure.

Study 2: Experimentally parsing out reading ability in established creativity measures. In May-Jun/2025, the study enrolled 1,100 participants across 140 high schools (grades 10-12) in 27 Brazilian States. The sample was drawn to represent all K–12 schools in the country. Recruited students participated in tablet-based creativity assessments, adapted from those conducted in Study 1.

We again proxy for SES by comparing students living in households with at most one full bathroom (*low-SES*) to those with two or more (*high-SES*). Based on this proxy, 681 study participants are low-SES, and 419, high-SES.

We adapted the same AUT and DAT tasks from Study 1 to be conducted on tablets in a representative sample of schools all over the country. Due to time limitations, we ran only two trials of the AUT (all subjects were prompted with a tire and a bottle, in that order), 30 seconds each, and one trial of the DAT, for 45 seconds.

All subjects received a seed word for the DAT ('ball'), which was not considered in their final score. Across all conditions, participants said their answers aloud, and enumerators recorded their responses.

We independently randomized whether each question prompt was read by the student (control group) or by the enumerator (treatment group). As such, in the control group, our measure of creativity may still reflect students' reading ability, while in the treatment group, it isolates creativity from reading skills. Across both treatment and control conditions, students responded verbally, and enumerators typed their responses; as a result, the creativity measure is independent of students' writing skills in both groups (see SI for all details).

In the AUT, the control group includes 559 participants, and treatment group, 541; in the DAT, the control group includes 546 participants, and the treatment group, 554. Participants' characteristics were balanced across control and treatment conditions (p-value of equality across all variables and conditions = 0.392).

Like Study 1, participants could refuse answering one or more trials, and response rates for each trial were statistically identical across SES (95.8% for low-SES and 96.5% for high-SES, p-value of difference = 0.442).

We compute p-values from Wald tests of differences in average test scores between high- and low-SES participants conditioning on whether participants read the prompt themselves or whether enumerators read them on their behalf. As in Study 1, we focus on a divergent thinking summary measure, stacking the dataset with participants' test scores in each trial (2 observations for AUT fluency and 1 for DAT) and clustering the analyses at the student level. We also control for age, gender, and location (= 1 if urban, 0 otherwise). Supplementary Materials showcase additional results: without controls, for each component of the summary measure, and using AUT originality instead of fluency.

Study 3: Experimentally parsing out reading ability in PISA measures. In Aug-Sep/2024, the study enrolled 515 participants across 43 high schools in 23 Brazilian States. The sample was drawn to represent all K–12 schools in the country. Recruited students participated in tablet-based creativity assessments, adapted from the PISA 2022 Creative Thinking Assessment.

We once again proxy for SES by comparing students living in households with one full bathroom (*low-SES*) to those with two or more (*high-SES*). Based on this proxy, 266 study participants are low-SES, and 249, high-SES.

In the experiment, we adapted two divergent thinking tasks from the PISA 2022 Creative Thinking Assessment framework. Each item assesses distinct ideation processes. The first item targets generating *diverse ideas*, emphasizing students' ability to think flexibly. The second item relates to generating *creative ideas*, focusing on the novelty and usefulness of an idea.

Specifically, the items were sourced from the Written Expression domain of the PISA 2022 Creative Thinking Assessment. In the first item, "Generate Diverse Ideas", high-school students are tasked with creating three distinct titles based on a surreal image displayed on a tablet, which features a large-scale book alongside various elements of the countryside, such as a field, tree, and wooden bench. They are given two minutes to complete this task. The second item, "Generate Creative Ideas", presents a dialogue between the sun and the earth in a comic strip format. This exercise consists of six dialogue boxes, whereby students must craft sentences for each character and communicate their responses to the enumerator within three minutes. These ideation processes embody cognitive skills essential for creative thinking in educational settings (17, 32, 34).

Responses were evaluated based on PISA's criteria for measuring creativity levels. Each item is associated with a specific difficulty level that determines the maximum score attainable in that item. A response can be rated as full credit, partial credit, or no credit. We resorted to two approaches to scoring students' responses. First, generative AI (GenAI), providing ChatGPT with item-specific prompts that include the evaluation criteria, classification parameters, and examples of responses associated with each credit score derived from the PISA materials. This approach leverages GenAI's capacity to analyze and interpret data, allowing us to evaluate student responses systematically (see SI for all details). Second, we had research assistants hand-code responses following the same prompts (see SI for all details). In both cases, our analysis focuses on the percentage of responses rated as *full credit*, based on the PISA scoring guidelines.

We once again independently randomized whether each question prompt was read by the student (control group) or by the enumerator (treatment group). Consistent with Study 2, students responded verbally across both treatment and control conditions, and enumerators typed their responses (see SI for all details).

In the first item, the control group includes 264 participants, and treatment group, 251; in the second item, the control group includes 243 participants, and the treatment group, 272. Participants' characteristics were balanced across control and treatment conditions (p-value of equality across all variables and conditions = 0.355).

As before, participants could refuse answering one or more trials; response rates for each trial were statistically identical across SES (80.8% for low-SES and 81.3% for high-SES, p-value of difference = 0.857).

We compute p-values from Wald tests of differences in average test scores between high- and low-SES participants conditioning on whether participants read the prompt themselves or whether enumerators read them on their behalf. As in Studies 1 and 2, we focus on a divergent thinking summary measure, stacking the dataset with participants' test scores in each trial (one for each task) and clustering the analyses at the student level. We also control for age, gender, and location, in line with Study 2. In the main text, we focus on hand-coded scores. Supplementary Materials present results without controls, for each component of the summary measure, and using GenAI scores instead of hand-coded ones.

Data Availability

All data are deposited at Open Science Framework (https://osf.io/d7hm8/).

Code Availability

A complete replication package is deposited at Open Science Framework (https://osf.io/d7hm8/).

Acknowledgements

Data collection for Study 1 conducted in partnership with the Municipal Secretariat of Education of Rio de Janeiro. Data collection for Studies 2 and 3 conducted by Equidade.info, generously funded by the Stanford Lemann Center. We acknowledge helpful discussions with Ben Domingue, Marily Oppezzo and Dan Schwartz, and excellent research assistance by João Bigon, Lucas Garcia, Paulina Huang, Rodrigo Megale, Bianca Mesquita, Jônatas Ribeiro, Karine Roncete and Mavigson Silva. All remanining errors are ours.

Author Contributions

G.L. conceptualized and supervised the study and wrote the manuscript. L.L. and S.A. developed the code, analyzed the data, and wrote the manuscript. All authors contributed to reporting the findings.

Competing Interests

None of the authors have competing interests to disclose.

References

1. Mauricio Castillo-Vergara, Nicole Barrios Galleguillos, Laura Jofré Cuello, Alejandro Alvarez-Marin, and Christian Acuña-Opazo. Does socioeconomic status influence student creativity? *Thinking Skills and Creativity*, 29:142–152, 2018. ISSN 1871-1871. . URL https://www.sciencedirect.com/science/article/pii/S1871187118300361.
2. M.R. Sarsani. Socio-Economic Status and Performance on Creativity Tests. In *Encyclopedia of Creativity*, pages 360–363. Elsevier, 2011. ISBN 978-0-12-375038-9. . URL https://linkinghub.elsevier.com/retrieve/pii/B9780123750389001795.
3. Selcuk Acar, Harun Tadik, Recep Uysal, Danielle Myers, and Betul Inetas. Socio-economic status and creativity: A meta-analysis. *The Journal of Creative Behavior*, 57(1):138–172, 2023.
4. Zhitian Skylor Zhang, Linda Hoxha, Abdullah Aljughaiman, Aliriza Arënliu, Maria P. Gomez-Arizaga, Sule Gucyeter, Irina Ponomareva, Jiannong Shi, Paula Irueste, Silke Rogl, Miguelina Nunez, and Albert Ziegler. Social Environmental Factors and Personal Motivational Factors Associated with Creative Achievement: A Cross-Cultural Perspective. *The Journal of Creative Behavior*, 55(2):410–432, 2021. ISSN 2162-6057. . URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jocb.463.
5. Karina Hendrie Kupczyszyn, Vanessa Arán Filippetti, and Laura Oros. Socioeconomic status effects on children's creativity: The mediating role of executive functions. *Thinking Skills and Creativity*, 51:101437, 2024. ISSN 1871-1871. . URL https://www.sciencedirect.com/science/article/pii/S1871187123002055.
6. Zhongyao Lu, Yaqiong Ding, and Yanhong Nie. How does family socioeconomic status affect creativity? the role of creative self-efficacy and critical thinking disposition. *Current Psychology*, 43(6):5674–5681, 2024.
7. David Yun Dai, Xiaoyuan Tan, Deepti Marathe, Anna Valtcheva, Robert M Pruzek, and Jiliang Shen. Influences of social and educational environments on creativity during adolescence: Does ses matter? *Creativity Research Journal*, 24(2-3):191–199, 2012.
8. Sofie Dahlman, Per Bäckström, Gunilla Bohlin, and Örjan Frans. Cognitive abilities of street children: Low-ses bolivian boys with and without experience of living in the street. *Child Neuropsychology*, 19(5):540–556, 2013.
9. Janet Rafner, Roger E Beaty, James C Kaufman, Todd Lubart, and Jacob Sherson. Creativity in the age of generative ai. *Nature Human Behaviour*, 7(11):1836–1838, 2023.
10. OECD. *Student performance in creative thinking.* OECD Publishing, Paris, 2024.
11. OECD. Pisa 2022 creative thinking, 2022. URL https://www.oecd.org/en/topics/sub-issues/creative-thinking/pisa-2022-creative-thinking.html. Accessed: 2024-10-10.
12. Anna Gomez, Elena Pecina, Sara Villanueva, and Tanya Huber. The undeniable relationship between reading comprehension and mathematics performance. *Issues in Educational Research*, 30(4):1329–54, 2020.
13. Gregory Camilli. Test fairness. In *Educational measurement (4th ed.)*, pages 221–256. American Council on Education/Praeger, 2006.
14. J. B. Carroll. *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. Cambridge University Press, 1993.
15. Jonathan Wai and Matt Brown. Developmental histories facilitating the emergence of creative scientific expertise: The role of developed cognitive talents, education, and social and cultural contexts. *Front. Psychol.*, 12, 2021. . URL https://doi.org/10.3389/fpsyg.2021.716529.
16. Janika Saretzki, Boris Forthmann, and Mathias Benedek. A systematic quantitative review of divergent thinking assessments. *Psychology of Aesthetics, Creativity, and the Arts*, 2024. .
17. Jay A Olson, Johnny Nahas, Denis Chmoulevitch, Simon J Cropper, and Margaret E Webb. Naming unrelated words predicts creativity. *Proceedings of the National Academy of Sciences*, 118(25):e2022340118, 2021.
18. Paul J. Silvia, Beate P. Winterstein, John T. Willse, Christopher M. Barona, Joshua T. Cram, Karl I. Hess, Jenna L. Martinez, and Crystal A. Richard. Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts*, 2(2):68–85, 2008. .
19. Andreas Fink, Thomas Reim, Mathias Benedek, and Roland H. Grabner. The Effects of a Verbal and a Figural Creativity Training on Different Facets of Creative Potential. *The Journal of Creative Behavior*, 54(3):676–685, 2020. . URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jocb.402.
20. Douglas Bors. The factor-analytic approach to intelligence is alive and well: A review of Carroll, J.B. (1993). Human cognitive abilities: A survey of factor-analytic studies. *Canadian Journal of Experimental Psychology*, 47:763–766, 1993.
21. J. P. Guilford. Creativity. *American Psychologist*, 5(9):444–454, 1950. . URL https://doi.org/10.1037/h0063487.
22. Mark A Runco and Selcuk Acar. Divergent thinking as an indicator of creative potential. *Creativity research journal*, 24(1):66–75, 2012.
23. Young J Juhn, Timothy J Beebe, Dawn M Finnie, Jeff Sloan, Philip H Wheeler, Barbara Yawn, and Arthur R Williams. Development and initial testing of a new socioeconomic status measure based on housing data. *Journal of Urban Health*, 88:933–944, 2011.
24. Igor Ferreira. Censo 2022: Rede de esgoto alcança 62,5% da população, mas desigualdades regionais e por cor e raça persistem | Agência de Notícias, 2024. URL https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/39237-censo-2022-rede-de-esgoto-alcanca-62-5-da-populacao-mas-desigualdades-regionais-e-por-cor-e-raca
25. J. D. Yeatman, K. A. Tang, P. M. Donnelly, M. Yablonski, M. Ramamurthy, I. I. Karipidis, S. Caffarra, M. E. Takada, K. Kanopka, M. Ben-Shachar, and B. W. Domingue. Rapid online assessment of reading ability. *Scientific Reports*, 11(1):6396, 2021. . URL https://doi.org/10.1038/s41598-021-85907-x.
26. J. D. Yeatman, J. E. Tran, A. Burkhardt, W. A. Ma, J. Mitchell, M. Yablonski, and A. Richie-Halford. Development and validation of a rapid online sentence reading efficiency assessment, December 1 2023. URL https://doi.org/10.31219/osf.io/u3mjz. Preprint.
27. J. D. Yeatman, C. Townley-Flores, K. Wentzlof, W. A. Ma, J. M. Siebert, M. Fuentes-Jimenez, A. Saavedra, T. S. Murray, K. Bhat, M. Ramamurthy, and ROAR Developer Consortium. Rapid online assessment of reading (roar): Technical manual. Technical report, Stanford

GSE, November 5 2024.
28. W. A. Ma, A. Richie-Halford, A. Burkhardt, K. Kanopka, C. Chou, B. Domingue, and J. D. Yeatman. Roar-cat: Rapid online assessment of reading ability with computerized adaptive testing. *Working Paper*, 2023.
29. Karine Roncete, Lucas Klotz, Wanjing Ma, Emily Arteaga, Luciana Alves, Rebecca Chrispim, Danielle Diniz, Jason Yeatman, and Guilherme Lichand. Development and validation of a rapid online sentence reading efficiency assessment, 2024. URL https://www.researchsquare.com/article/rs-5516837/v1. Preprint.
30. Sachin Allums, Karine Roncete, and Guilherme Lichand. Generative AI can harm learning despite guardrails: Evidence from middle-school creativity. *mimeo*, 2025.
31. Peter Organisciak, Selcuk Acar, Denis Dumas, and Kelly Berthiaume. Beyond semantic distance: Automated scoring of divergent thinking greatly improves with large language models. *Thinking Skills and Creativity*, 49:101356, September 2023. ISSN 1871-1871. . URL https://www.sciencedirect.com/science/article/pii/S1871187123001256.
32. OECD. *Measuring creative thinking*. OECD Publishing, Paris, 2024.
33. Joy Paul Guilford. The structure of intellect. *Psychological bulletin*, 53(4):267, 1956.
34. Brian J Lucas and Loran F Nordgren. The creative cliff illusion. *Proceedings of the National Academy of Sciences*, 117(33):19830–19836, 2020.

Figures

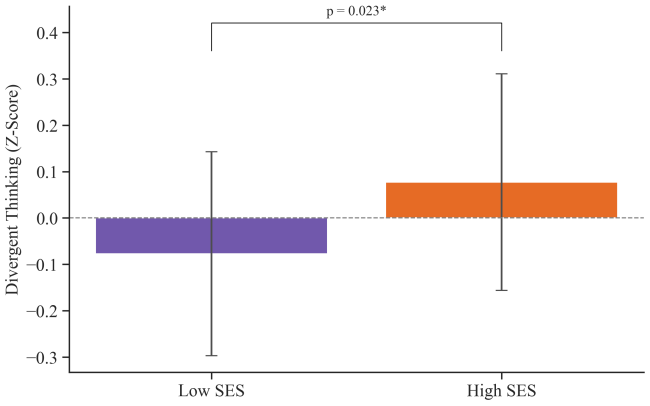

Fig. 1. SES creativity gap in divergent thinking summary measure

Notes: Summary measure of two trials of the Alternative Uses Task (AUT) and one trial of the Divergent Association Task (DAT), stacking datasets (i.e., each observation is a student × trial) while standardizing the outcome by the mean and standard deviation of the whole sample within each trial. Low-SES participants report to live in households with at most 1 full bathroom; high-SES ones, with 2 or more full bathrooms. N = 523 (366 low-SES and 157 high-SES participants). P-value of a Wald test of difference in means, clustering standard errors at the participant level.

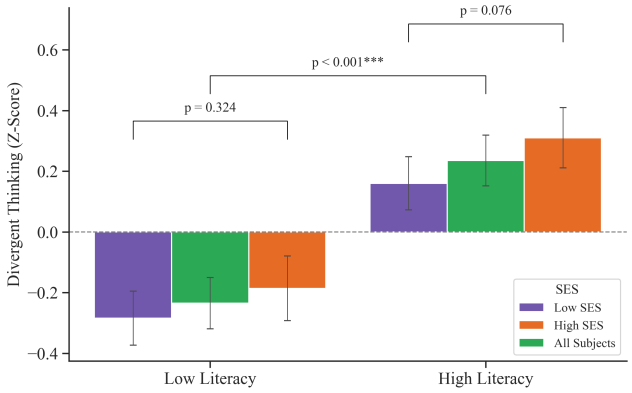

Fig. 2. SES creativity gaps, by reading ability

Notes: Summary measure of two trials of the Alternative Uses Task (AUT) and one trial of the Divergent Association Task (DAT), stacking datasets (i.e., each observation is a student × trial) while standardizing the outcome by the mean and standard deviation of the whole sample. Low-SES participants report to live in households with at most 1 full bathroom; high-SES ones, with 2 or more full bathrooms. High (low) literacy participants are those who do (not) read at grade-5 level based on ROAR-SRE (see 29). N = 523 (366 low-SES and 157 high-SES participants). P-value of Wald tests of differences in means within reading ability level, clustering standard errors at the participant level.

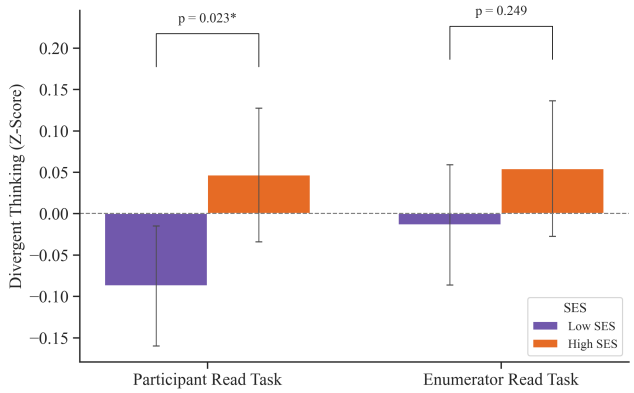

Fig. 3. SES creativity gaps in standard measures, by experimental condition

Notes: Summary measure of two trials of the Alternative Uses Task (AUT) and one trial of the Divergent Association Task (DAT), stacking datasets (i.e., each observation is a student × trial) while standardizing the outcome by the mean and standard deviation of the control group (in which participants read the prompt) within each trial. Low-SES participants report to live in households with at most 1 full bathroom; high-SES ones, with 2 or more full bathrooms. N = 1,100 (681 low-SES and 419 high-SES participants). P-value of Wald tests of differences in means within each experimental condition, clustering standard errors at the participant level.

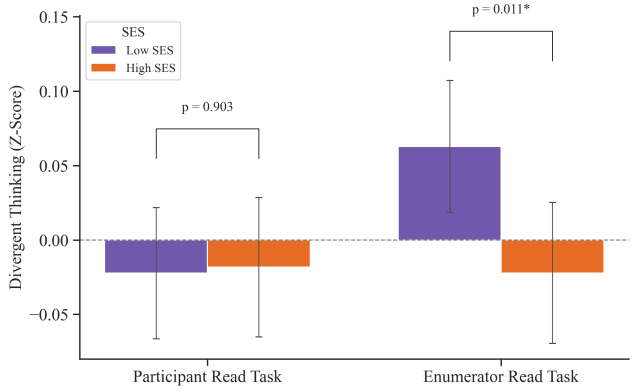

Fig. 4. SES creativity gaps in PISA-like items, by experimental condition

Notes: Summary measure of two items adapted from the PISA 2022 Creative Thinking Assessment, stacking datasets (i.e., each observation is a student × trial) while standardizing the outcome by the mean and standard deviation of the control group (in which participants read the prompt) within each trial. Low-SES participants report to live in households with at most 1 full bathroom; high-SES ones, with 2 or more full bathrooms. N = 515 (266 low-SES and 249 high-SES participants). P-value of Wald tests of differences in means within each experimental condition, clustering standard errors at the participant level.