PISA

Seven Questions about Creativity and Creative Thinking

What Do PISA 2022 Data Tell Us?

PISA

Seven Questions about Creativity and Creative Thinking

WHAT DO PISA 2022 DATA TELL US?

OECD
BETTER POLICIES FOR BETTER LIVES

Please cite this publication as:
OECD (2025), *Seven Questions about Creativity and Creative Thinking: What Do PISA 2022 Data Tell Us?*, PISA, OECD Publishing, Paris, https://doi.org/10.1787/0aa52128-en.

ISBN 978-92-64-64690-2 (print)
ISBN 978-92-64-94974-4 (PDF)
ISBN 978-92-64-91969-3 (HTML)

PISA
ISSN 1990-8539 (print)
ISSN 1996-3777 (online)

Photo credits: Illustrations © Fonzy Nils

Corrigenda to OECD publications may be found at: https://www.oecd.org/en/publications/support/corrigenda.html.

Foreword

This report summarises the results of a year-long collaboration between the OECD Secretariat and researchers in 14 countries (16 country-language contexts) focused on rescoring a sub-sample of student responses to select tasks from the PISA 2022 Creative Thinking assessment. The purpose of the "PISA CT Rescoring project" was to conduct further research on the rich data contained in student responses to the open-ended PISA creative thinking items, as well as to pilot alternative scoring methods that can be used to measure creative thinking in these types of task and that can serve as examples for national and international assessment approaches.

The PISA CT Rescoring project involved over 100 judges that applied three different scoring methods – a holistic judgment approach, a criteria-based analytical approach, and a conceptual-knowledge mapping – to nearly 38,000 unique responses from 14,830 students that took the PISA 2022 test in 16 country-language contexts. These three approaches conceptualise and prioritise creative value in different ways to each other and to the international scoring rules applied in the PISA 2022 CT assessment. In particular, the new scoring methods enable the *most creative* responses within this sample to be identified, while the scoring method applied in PISA 2022 primarily focused on distinguishing creative (original) from non-creative responses. As such, the PISA CT Rescoring project does not aim to replicate the PISA 2022 results but rather to explore alternative ways of evaluating open-ended tasks that ask students to think creatively.

The first results from this project were described in an earlier report published by the OECD titled *Creative Minds in Action,* which showcased genuine student responses to the PISA 2022 creative thinking tasks that were judged to be amongst the most creative in the sample across country-language contexts.

In this report, *Seven Questions About Creativity and Creative Thinking*, we analyse the results from the PISA CT Rescoring project, together with the PISA 2022 test and questionnaire data. These data allow us to investigate key questions related to measuring creative thinking across different socio-cultural contexts. While cultural differences exist in judges' subjective evaluations of students work, overall, judges across the country-language groups involved in the Rescoring project all assigned strong importance to the essential criteria defining creativity – that is, appropriateness, originality and value. The rescored data also allow us to examine the extent to which key findings from the report are replicated when focusing on different dimensions of creative value. For example, the rescored data show that girls performed better boys, on average, even when changing how their responses are evaluated (e.g. considering the overall creative value of their responses). Socio-economic differences in performance also remain marked across the different scoring methods, with disadvantaged students particularly struggling in tasks that required longer written responses. These and other findings described in this report confirm the value of the data generated by the PISA test, and indicate that further research using the PISA data can uncover other important lessons for education policy and practitioners that wish to nurture creative talent.

Acknowledgements

This report is the product of a collaborative effort between countries and economies that participated in the PISA 2022 creative thinking assessment, the OECD PISA Secretariat, and research teams across country-language groups. This report was drafted by Natalie Foster and Said Ettejjari, with support from Sofia Bertolaja, under the guidance of Andreas Schleicher, Yuri Belfali and Mario Piacentini. Sophie Limoges co-ordinated the production for publication, Della Shin designed and laid out the report, and Jessica Holmes provided editorial support.

The OECD gratefully acknowledges and sincerely thanks the many individuals involved in the PISA CT Rescoring project for their conceptual and methodological contributions. We particularly thank: Maud Besançon, Anaelle Camarda and Todd Lubart as international project coordinators; Anaelle Camarda and Mathieu Cassotti for developing the C-K Tree schema used to score two tasks; the LEGO Foundation for their generous support; the lead researchers in each country-language team; and all of the judges below.

Belgium (French community)
Baptiste Barbot (Lead)
Shannon Crappe
Clara Godart
Eva Hardy
Yaroslava Mazurenko
Jean Urbaniak

Brazil
Ricardo Primi (Lead)
Solange M. Wechsler (Lead)
Tatiana de Cassia Nakano
Beatriz Chrispim
Natalia de Oliveira Santos
Ellen Carolina de Paula Faria
Paloma do Nascimento Vilvert
Asheley dos Santos da Paixão
Vivian Nascimento da Silva
Cibele Oliveira
João Vítor Sarôa Brandine

Canada (English provinces)
Linxuan (Marina) Wang (Lead)
Yong Zeng (Lead)
Charlotte Alexander
Rupinder Kaur
Mohan Qiao
Sierra Robinson
Hamed Shirazi
Chang Su
Jiami Yang
Morteza Zangeneh Soroush

Canada (French provinces)
Xavier Robichaud (Lead)
Samuel Blanchard
Francis Bourgoin
Nicholas Derby
Caitlin Furlong
Dominic Manuel

Chile
David Preiss (Lead)
Benjamín Carmona
Catalina Reszczynski

Colombia
Manuel Amado (Lead)
Yuly Paola Martínez Sánchez (Lead)
Daisy Pilar Ávila Torres
Martha Jeanet Castillo Ballén
Julián Andrés Penagos Mora
Raúl David Torres Reyes
Betsy Yamil Vargas Romero

France
Maud Besançon (Lead)
Anaelle Camarda (Lead)
Todd Lubart (Lead)
Amaya Allende Santa Cruz
Jonathan Bique
Aline Chaplain
Chandra David
Morgane Facon
Louann Follezou
Juliette Jacquin
Emeline Le
Eliott Mahe
Pauline Perez
Noélyne Rouillé

Germany
Boris Forthmann (Lead)
Mathias Benedek
Robin Bosse
Leonie Cohn
Jennifer Diedrich
Benjamin Goecke
Jule Neuß
Selina Weiss

Greece
Dimitrios Zbainos (Lead)
Dimitrios Karatzanos
Charis Sagia

Hong Kong (China)
Weiping Hu (Lead)
Yangping Li
Xue Li
Huizhi Bai

Italy
Sergio Agnoli (Lead)
Giovanni Emanuele Corazza (Lead)
Lorenzo Campidelli
Angela Faiella
Kirill Miroshnik
Stefano Rini
Cecilia Segatta

Portugal
Irene Cadime (Lead)
Leandro Silva Almeida (Lead)
Ana Isabel Almeida
Paula Batista
Albertina Ferreira
Catarina Figueiredo
Filipa Silva
Stephanie Silva
Felipe Zamana

Saudi Arabia
Abdullah Aljughaiman (Lead)
Rudaynah Mohammed Alawad Alfadhli
Malak Adbulaziz Ali Alabdullatif
Riham Abdulrahman Aljafari
Qatnah Alhumaidi Al-Shammari
Qamrah Alsubaie
Norah Fareed Almulhim
Norah Saad Al-Dossary

Slovenia
Mojca Juriševič (Lead)
Karin Bakračevič
Ana Mlekuž
Neža Podlogar
Klaudija Šterman Ivančič

Korea
Kyung-hwa Lee (Lead)
Hyunwoo Kim (Lead)
Jusung Jun
Songhyun Park
Soojin Park
Hee-ra Bae
Hye-jin Yang
Jinkyung Yoon

Table of Contents

Section 2. Unpacking cross-cultural differences in creativity and creative thinking

Section 3. Variation in creative thinking performance across student groups

Annexes

Tables

Figures

Boxes

Executive summary

What makes "creative ideas" creative?

Creative ideas tend to be highly appropriate, highly original and highly valuable across country-language contexts. However, judges' appropriateness, originality and value scores don't fully explain their overall evaluations of the creativity of students' responses, accounting for around 66% of the variation in holistic creativity scores on average. The relative importance of appropriateness, originality and value also varies across task types, though they remain important in all contexts. Factors like task demands, constraints and difficulty affect how much each criteria determines the overall creativity of a response.

Which tasks led to more creative ideas on the PISA test?

Some of the tasks in the PISA 2022 Creative Thinking test asked students to come up with two or three ideas for a given scenario that were as different as possible. Although the task instructions did not explicitly instruct students to think of creative ideas in these tasks, around 1 in 2 students managed to come up with at least one creative idea in their response. In fact, students were more successful in thinking of creative ideas in the divergent thinking tasks than when asked to generate a single original idea, with only 1 in 4 student responses evaluated as creative in those tasks. Divergent thinking tasks can thus encourage creative idea generation and help students practice their ideation skills. Students might also need support in evaluating and selecting good ideas (i.e. convergent thinking processes) following divergent thinking.

Is creative thinking domain-specific or domain-general – or both?

The PISA 2022 Creative Thinking test included tasks situated in different domain contexts, some of which focused more on imagination (e.g. writing a story) while others focused more on solving social or scientific problems. Of the students in the PISA CT Rescoring project sample who completed at least three different tasks, only 8% on average were high-scoring (compared to their country-language peers) in three or more tasks. This finding suggests that it is quite difficult for students to be creative across all domains.

Students' gender, socio-economic status, and mathematics and reading performance were all significantly associated with scoring highly across multiple tasks. Creative all-rounders also reported stronger attitudes and beliefs related to creativity. However, a range of other domain-specific and domain-general factors likely also contribute to supporting creative performance across different domains.

Are there cross-cultural differences in creativity and creative thinking?

Cultural norms and expectations can influence students' capacity to engage in creative work and how people evaluate what is creative or not. Unsurprisingly, across all country-language groups in the study, students found in relatively easier to come up with appropriate ideas across tasks than original and valuable ideas. Students also consistently demonstrated a relative weakness in coming up with original ideas, which was particularly evident in some countries (like Chile or Greece). It may be that judges have higher expectations about what makes an idea original in the types of everyday imaginative and problem-solving tasks in the PISA test.

Despite students' weaknesses in coming up with original ideas, originality scores explained the largest share of the unique variation in judges' holistic creativity scores, on average across country-language groups – though not in all country-language groups. In Greece, Colombia, Chile, Korea, appropriateness scores instead explained the largest share of the unique variation in creativity scores. While value scores often explained more of the unique variation in judges' scores than appropriateness scores, they rarely explained more than originality scores (only in Greece, Korea and Italy).

Patterns in idea flexibility and fixations in problem-solving tasks

In two of the problem-solving tasks included in the Rescoring study, students tended to think of similar ideas – both within and across country-language groups. In both tasks, 5 most common idea types accounted for over 60% of all student ideas, respectively. There was greater cross-cultural variability in students' ideas in the harder scientific problem-solving task than the easier social problem-solving task.

Students were also far more likely to ignore the task constraints in the more difficult task. Students who submitted ideas that followed relatively common themes also scored higher, in terms of overall idea quality, than students who suggested ideas in more uncommon idea categories. These creative thinkers "inside the box" were more likely to be girls and socio-economically advantaged students.

Are girls really better than boys in creative thinking?

Girls consistently scored higher than boys in the PISA 2022 creative thinking tasks – across country-language groups, task types and scoring method – suggesting it is a relatively stable finding that is not driven by any single scoring method, task feature, or sample anomaly. Girls' greater engagement with the PISA test, in general, may go some way to explain this result.

The only domain in which girls did not show a large performance advantage over boys in creative thinking is in scientific problem solving. Despite still outscoring boys in scientific problem-solving tasks, girls showed higher levels of relative disengagement than boys with the scientific problem-solving tasks and they reported far weaker creative self-efficacy in scientific or invention tasks compared to boys (despite reporting overall greater creative self-efficacy).

Does socio-economic disadvantage hinder creative thinking?

Student socio-economic status was significantly associated with their creativity scores – particularly in countries in the study with the largest within-country inequalities in socio-economic status (e.g. Colombia, Brazil and Portugal). Advantaged students scored significantly higher than disadvantaged students, both overall and in idea appropriateness, originality and value. Disadvantaged students particularly struggled relative to advantaged students in tasks asking them to generate original ideas, tasks with greater writing demands, and problem-solving tasks.

Differences in reading and mathematics proficiency between advantaged and disadvantaged student groups largely – but not totally – account for their performance differences in creative thinking. Self-efficacy beliefs also effectively moderate students' success in creative thinking tasks in all country-language groups, and especially in scientific problem-solving tasks (on average).

Introduction

Every three years, a sample of 15-year-old students around the world sits an assessment, known as PISA (Programme for International Student Assessment), that aims to measure how well their education system has prepared them for life after compulsory schooling. In 2022, for the first time, the PISA assessment included a test of creative thinking that was administered in 64 participating countries and economies. As part of PISA 2022, students also completed a background questionnaire about themselves, their homes and schools, and their experiences at school.

The PISA 2022 report on creative thinking was released in June 2024 and summarised the international results from the assessment (OECD, 2024[1]). The results provide a comparison of the extent to which 15-year-old students in each participating country and economy can engage in the process of creative thinking to generate diverse ideas or to generate and improve upon original ideas. Key findings from the report include that a significant gender gap exists in creative thinking (girls consistently outperformed boys); that proficiency in reading, mathematics and science supports performance in creative thinking but, beyond a baseline proficiency in these core literacies, all students can excel in creative thinking; and that performance differences associated with socio-economic status persist in creative thinking, despite an overall weaker association than the one observed between student background and performance in the core domains. Some countries and economies also performed better than expected in the creative thinking assessment, with the report identifying features of these high-performing systems.

While many national centres and governments try to ensure that educators and educational researchers get constructive feedback based on PISA results, most of the key messages published in the PISA reports don't make it back to those working to support students in classrooms. What's more, understanding how creative ideas develop and manifest has been the subject of several decades of research – but the PISA 2022 assessment represents the first large-scale, internationally comparable data collection of its kind. These data can shed new light on how successful 15-year-olds are at creative thinking around the world, the different ways that creative success can be measured, the variation in performance observed across student groups, and whether there are cross-cultural differences in how creativity is understood and evaluated.

Data and results in this report

This report analyses a sample of the data that was collected during the PISA 2022 Creative Thinking (CT) assessment. These data – students' raw responses to a subset of the PISA 2022 CT items (see Insert) – were rescored as part of a research project (the "PISA CT Rescoring project") conducted by the OECD and collaborating research teams in 16 country-language groups to further explore the rich information in students' responses to the PISA test. The main goals of the PISA CT Rescoring project were twofold:

1. To investigate the feasibility of applying different scoring methods to the PISA 2022 data across diverse socio-cultural contexts (see Box A);
2. To analyse the rescored data sample and examine the extent to which key findings from the PISA 2022 report are replicated across different scoring methods.

The three scoring methods applied in the PISA CT Rescoring project conceptualise and prioritise creative value in different ways to each other and to the international scoring rules applied in the PISA 2022 CT assessment. In the PISA 2022 CT assessment, responses to creative thinking items were awarded full credit, partial credit (in some cases), or no credit. The scoring rules, described in the PISA 2022 CT assessment framework (OECD, 2023[2]), focused on evaluating students' capacity to generate "sufficiently" original or different ideas (depending on the item type) across contexts – or in other words, distinguishing sufficiently creative from non-creative responses. This approach was adopted to maximise the scalability and reliability of the human coding effort in each country/economy and to facilitate cross-cultural comparability by focusing on criteria that could be evaluated most objectively across diverse contexts (Foster and Piacentini, 2025[3]; Foster and Piacentini, forthcoming[4]). For example, statistical infrequency was adopted as a proxy for the originality of student responses. However, the trade-off of the PISA 2022 CT scoring approach was that it could not identify the most creative responses within each participating country/economy, nor did it evaluate the overall "creative value" of students' responses. In contrast, the alternative scoring methods applied in the PISA CT Rescoring project do provide different ways for the most creative responses in the sample to be identified. A first report using data from the PISA CT Rescoring project exploring students' ideas on the test was released earlier this year (OECD, 2025[5]).

Box A. PISA CT Rescoring project: Data and scoring approaches used in the study

Data and item sample

The PISA CT Rescoring project data sample included 37,849 unique responses from 14,830 students that sat the PISA 2022 CT test, sampled from 14 participating countries and economies (16 country-language groups) (Annex A1). 7 items from the full 32 test-item pool were included in the study (see Insert). A maximum of 1000 students from each country-language group were randomly sampled, with at least 300 students sampled per item (where possible).[1] Over 100 judges were involved in the rescoring effort, led by a team of experienced researchers in the field of creativity measurement in each of the participating country-language jurisdictions.

Scoring methods applied in the study

The three scoring methods applied to the PISA CT Rescoring project data sample were:

- The **Holistic Judgement method**: Valid task responses were given a single "creativity" score (score between 1-7 points), considering all relevant elements of their task response. The goal of this scoring method was to establish a relative ranking of responses within each country-language-task group.
- The **Criteria-Based method**: Each idea within a student task response was scored for its appropriateness (score between 0-2 points), originality and value (scores between 0-3 points, each). A cumulative idea sum score was also produced for each idea by aggregating the scores across these three criteria. In addition, responses to tasks requiring multiple ideas were scored for idea flexibility (one score per task response, between 0-2 points).
- The **Conceptual Knowledge Tree method**: Each idea within a student task response was mapped onto a branch and node of a conceptual knowledge tree that defines all possible solution path ideas in a problem space (Hatchuel and Weil, 2009[6]; Agogué et al., 2014[7]). This method was applied to two problem-solving tasks only (*Food Waste* and *The Exhibit* tasks – see Insert).

More information about the three scoring methods and the coding process that judges in each participating country-language group followed can be found in Annex A2. Information about the inter-rater reliability tests and results for each of the three methods can also be found in Annex A2.

Content of this report

This report analyses the results of the PISA CT Rescoring project and organises the findings into answers to 7 key questions about creativity and creative thinking (listed below). The report explores how these insights from the data might help educators and educational researchers to support the development of students' creative thinking.

The questions can be grouped into three main topic sections:

1. Understanding creativity and creative thinking processes;
2. Unpacking cross-cultural differences in creativity and creative thinking; and
3. Variation in creative thinking performance across student groups

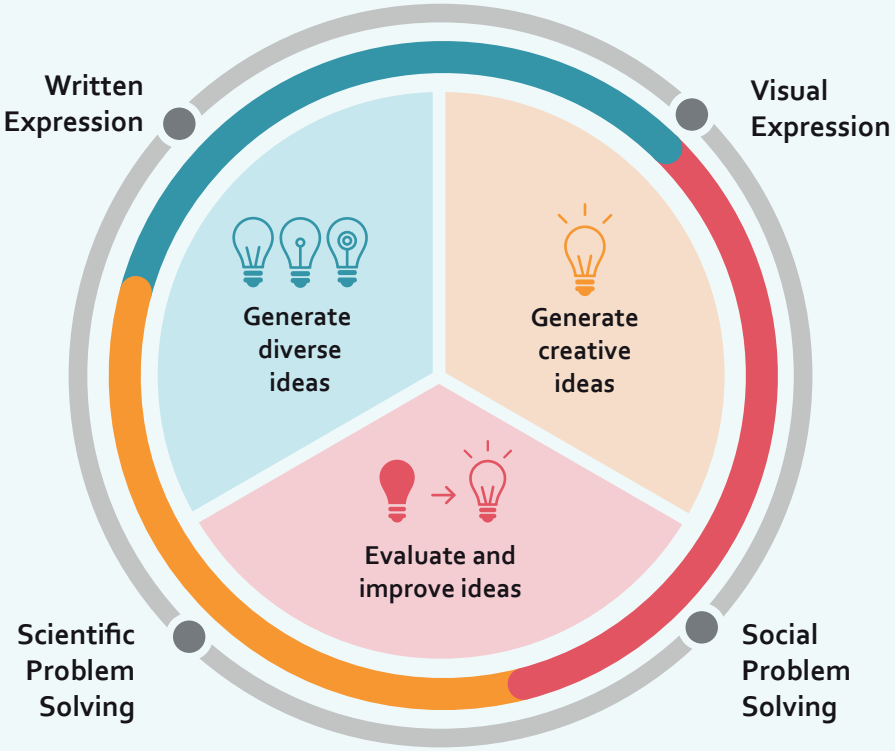
PISA 2022 tasks included in the PISA CT Rescoring Study

PISA 2022 student model: ideation processes and domain contexts

The PISA creative thinking construct was broken down into three ideation processes for measurement purposes in the PISA 2022 test (Figure A). These three ideation processes encompass both divergent-exploratory and convergent-integrative thinking processes and were classified as follows: (1) 'generate diverse ideas'; (2) 'generate creative ideas'; and (3) 'evaluate and improve ideas'. Every task on the PISA 2022 creative thinking test corresponded to one of these ideation processes.

Figure A. PISA 2022 creative thinking competency model

Ideation processes and domain contexts targeted in the PISA 2022 CT test



Source: (OECD, 2023[2])

The PISA 2022 CT test also situated tasks in four different domain contexts. Two contexts – written expression and visual expression – focused on students' capacity to express their ideas creatively, whereas the other two contexts – social problem solving and scientific problem solving – focused on students' capacity to apply creative thinking for a more functional purpose.

Every task in the test required students to submit a constructed response answer, either in the form of a written response (for the written expression and problem-solving tasks) or a graphic design (for the visual expression tasks).

Tasks included in the PISA CT Rescoring project

A balanced subset of 7 tasks from the wider 32 item pool were chosen for the PISA CT Rescoring study. These 7 tasks were chosen considering the following factors:

- Representation of the four domain contexts;
- Representation of divergent-exploratory and convergent-integrative thinking processes;
- Task position in the PISA 2022 test forms (to maximise the number of students in the sample that sat multiple items included in the study);
- Tasks already released into the public domain following the administration of PISA 2022.

Table A describes the balance of tasks included in the PISA CT Rescoring study by ideation process and domain context. Note that while no items classified as 'evaluate and improve ideas' items were included in the PISA CT Rescoring project, these tasks were similar in format and task demands to the 'generate creative ideas' items.

Table A. Balance of tasks included in the PISA CT Rescoring project

By ideation process and domain context

Tasks	Written expression	Visual expression	Social problem-solving	Scientific problem-solving	Total
Generate diverse ideas	1	0	1	2	4
Generate creative ideas	1	1	1	0	3
Total	2	1	2	2	7

Note: 'Generate diverse ideas' tasks required students to submit two or three different ideas. 'Generate creative ideas' tasks required students to submit only one original idea.

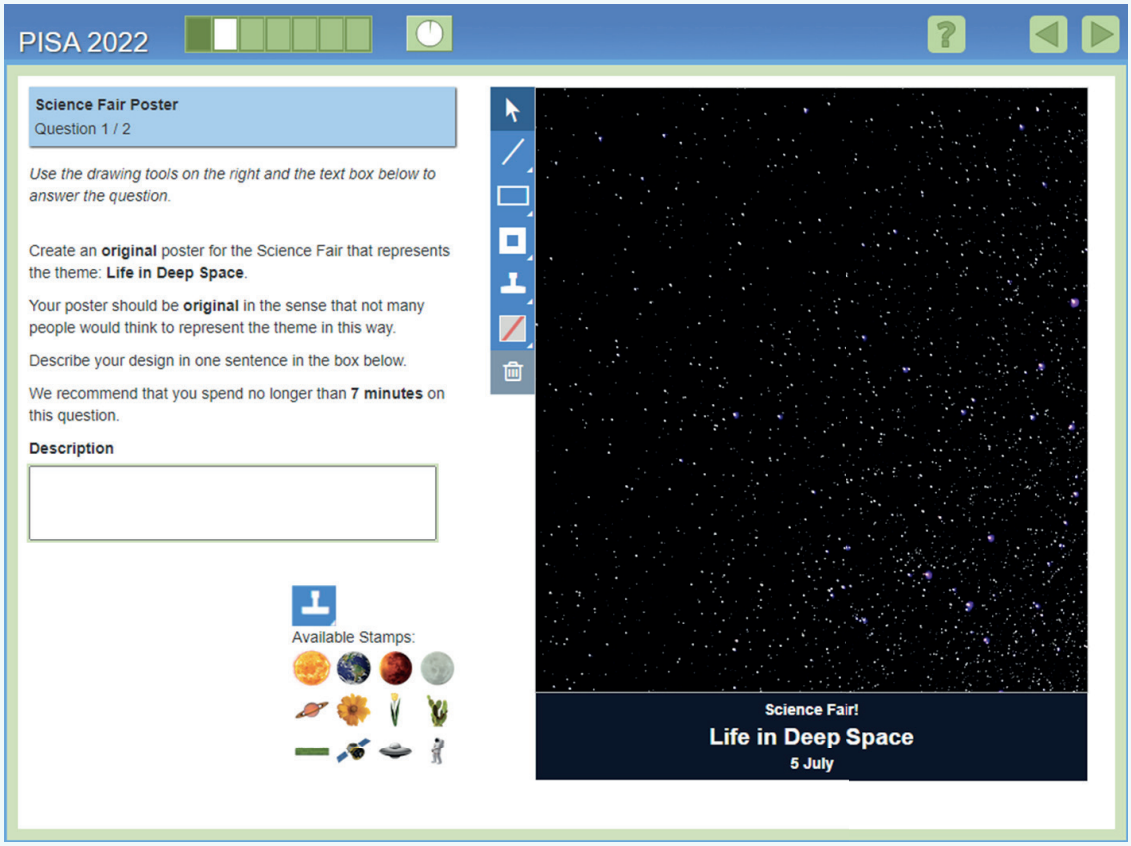
Each of the 7 tasks included in the PISA CT Rescoring study are shown below. For an in-depth exploration of genuine student responses to these 7 tasks, including some of the most creative responses in the project sample, see OECD (2025[5]).

Task 1 – Science Fair Poster

- Ideation process: Generate creative ideas
- Domain context: Visual expression

In the first task of the *Science Fair Poster* unit, students were asked to create an original poster for a school science fair that focused on the theme of "Life in Deep Space" (Figure B). The digital tool provided to students allowed them to drag and drop a selection of existing stamps and shapes in the workspace, as well as manipulate the colours of shapes and lines. Students were also allowed to provide a brief description of their posters to help judges interpret the meaning behind their designs.

Figure B. Science Fair Poster task



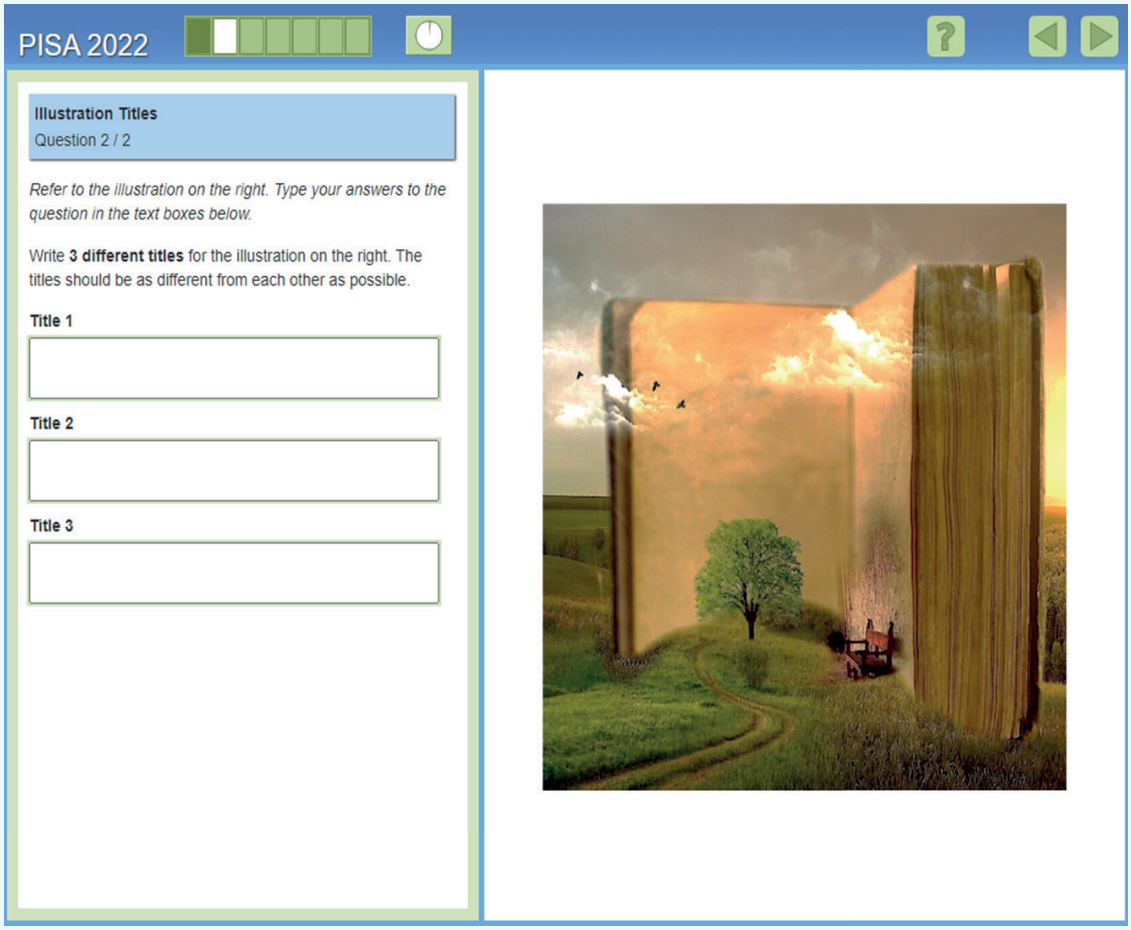
Note: This task is the first of two tasks in the *Science Fair Poster* unit.
Source: https://www.oecd.org/en/about/programmes/pisa/pisa-2022-creative-thinking-test-questions.html

Task 2 – Illustration Titles

- Ideation process: Generate diverse ideas
- Domain context: Written expression

For the second task in the *Illustration Titles* unit, students were asked to suggest creative title ideas for an abstract illustration showing a large book in a fantasy countryside landscape (Figure C). Students were instructed to think of three title ideas that were as different as possible to each other.

Figure C. Illustration Titles task



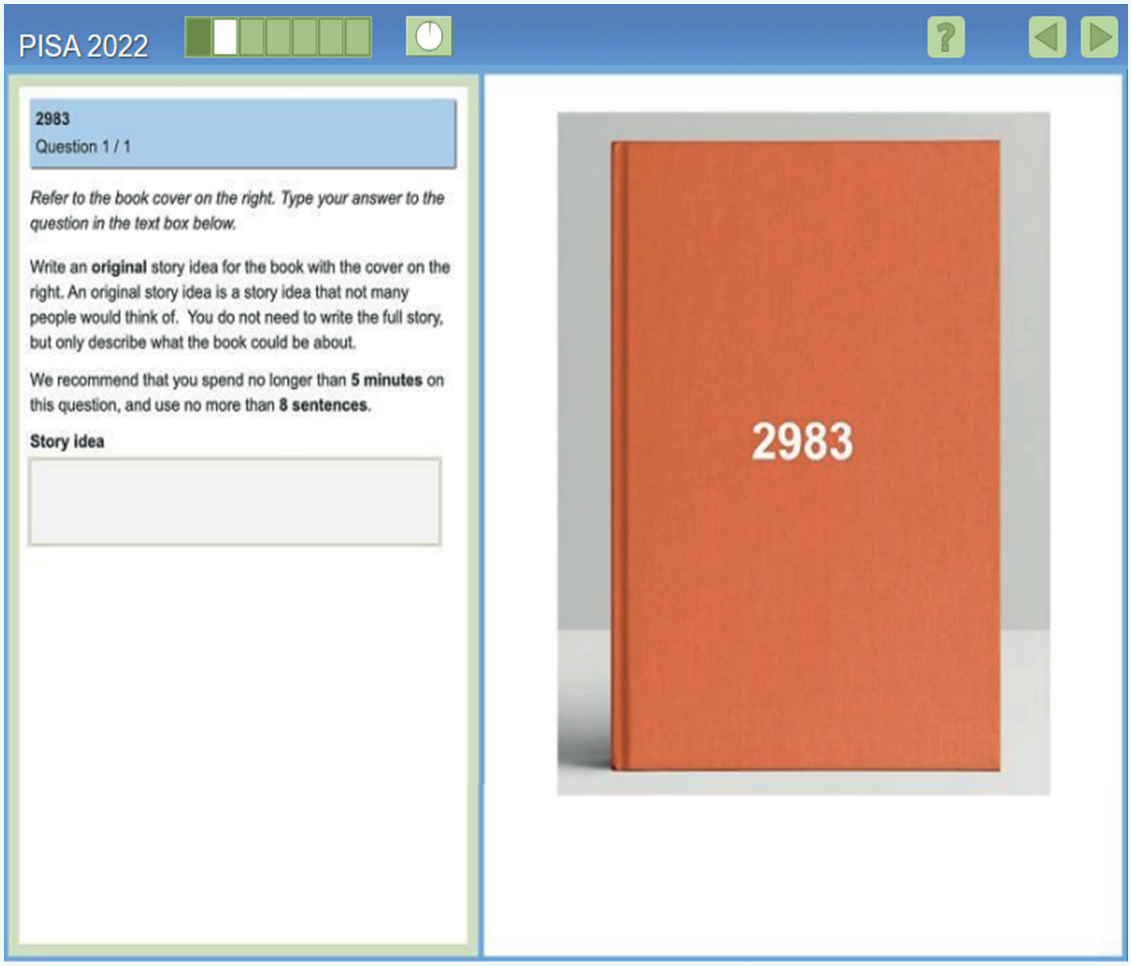
Note: This task is the second of two tasks in the *Illustration Titles* unit.
Source: https://www.oecd.org/en/about/programmes/pisa/pisa-2022-creative-thinking-test-questions.html

Task 3 – 2983

- Ideation process: Generate creative ideas
- Domain context: Written expression

In the written expression task *2983*, students were asked to think of story idea for a book whose cover displays only the 4-digit number 2983 (Figure D). Students were asked to briefly describe what the story could be about and to focus on making their story idea as original as possible.

Figure D. 2983 task



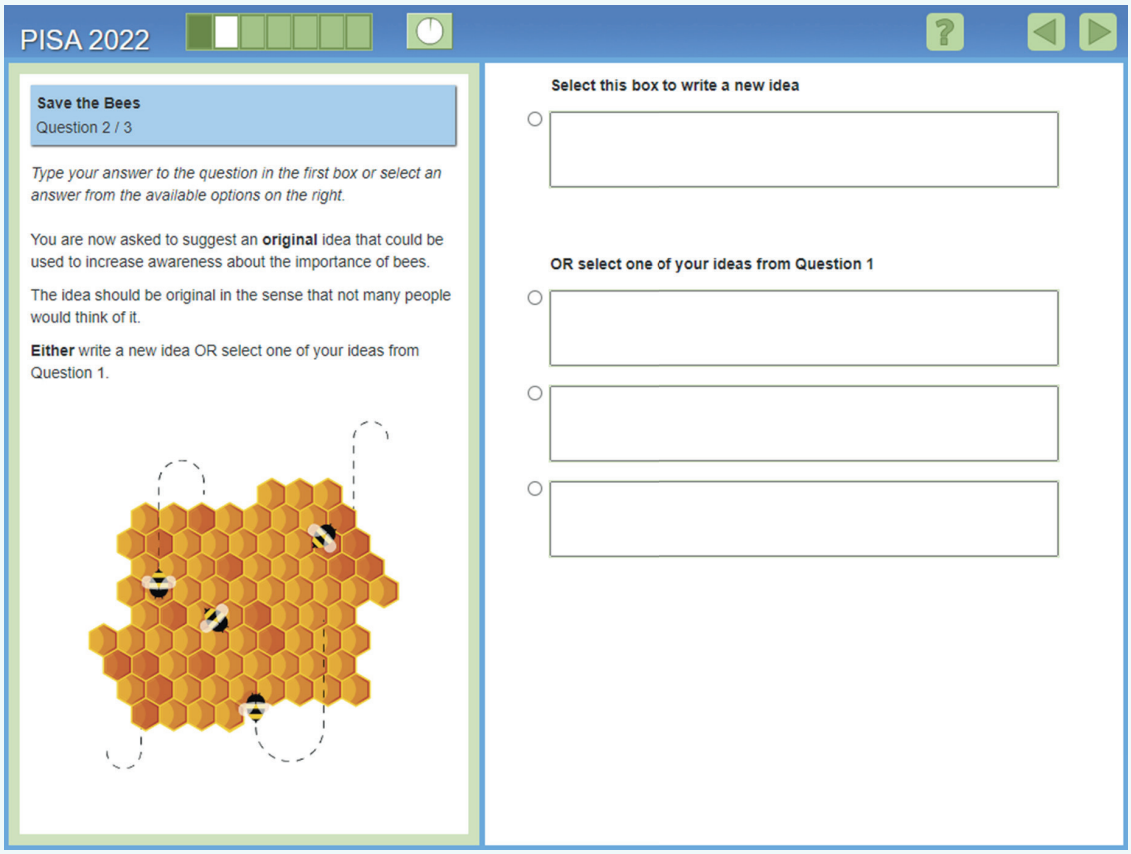
Note: This task was the only task in the *2983* unit.
Source: https://www.oecd.org/en/about/programmes/pisa/pisa-2022-creative-thinking-test-questions.html

Task 4 – Save the Bees

- Ideation process: Generate creative ideas
- Domain context: Social problem solving

In the *Save the Bees* unit, students are told that the "Save the Bees" club at their school are trying to increase awareness about the important role that bees play in local ecosystems. In the first task of the unit, students were asked to suggest three different ideas for how the club could raise awareness about the importance of bees in their school. In this second task (Figure E), students were asked to suggest one original solution for achieving the same goal. Students were given the choice of providing a completely new idea or choosing one of the ideas they came up with in the previous task.

Figure E. Save the Bees task



Note: This task is the second of three tasks in the *Save the Bees* unit. Students could write a new response or choose to submit one of their ideas from the previous task.

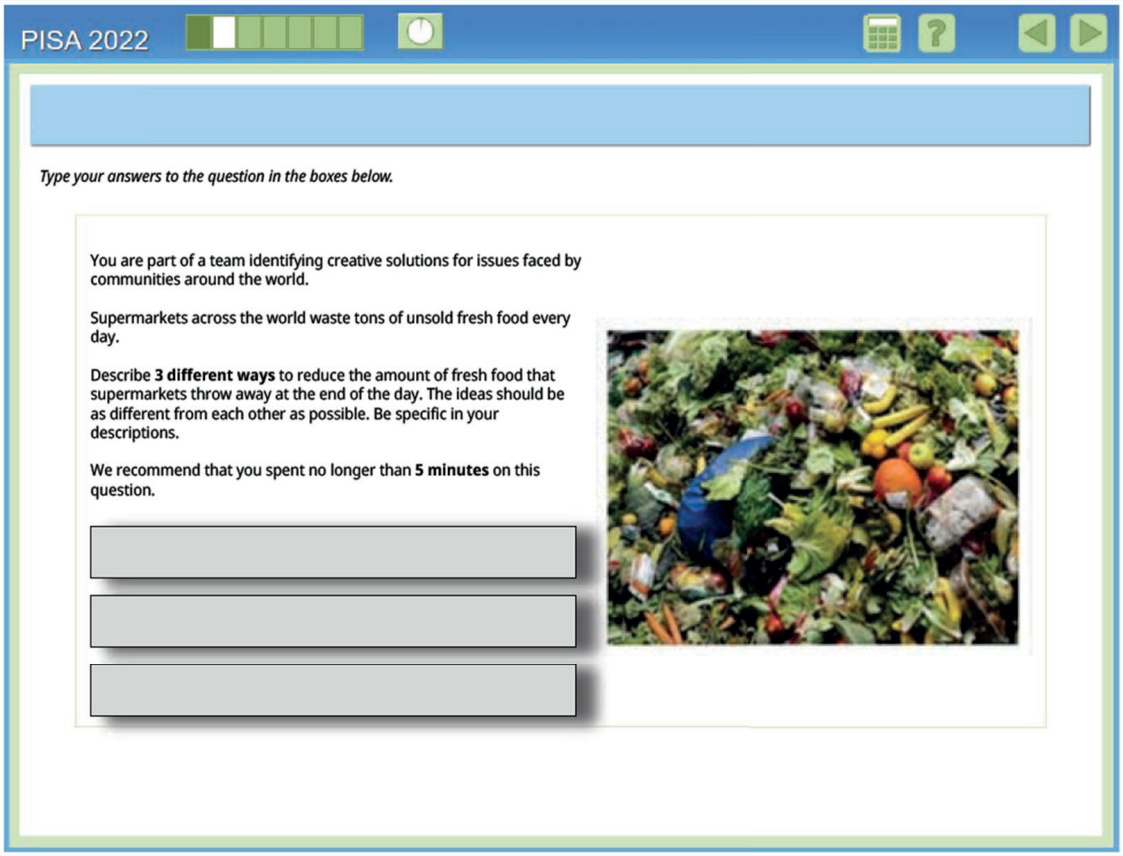
Source: https://www.oecd.org/en/about/programmes/pisa/pisa-2022-creative-thinking-test-questions.html

Task 5 – Food Waste

- Ideation process: Generate diverse ideas
- Domain context: Social problem solving

In the *Food Waste* unit, students were presented with the problem scenario that supermarkets across the world waste tons of unsold fresh food every day. Students were then asked to suggest three different ways that supermarkets could reduce the amount of fresh food that they throw away at the end of the day (Figure F).

Figure F. Food Waste task



Note: This task was the only task in the *Food Waste* unit.
Source: PISA 2022 Creative Thinking test.

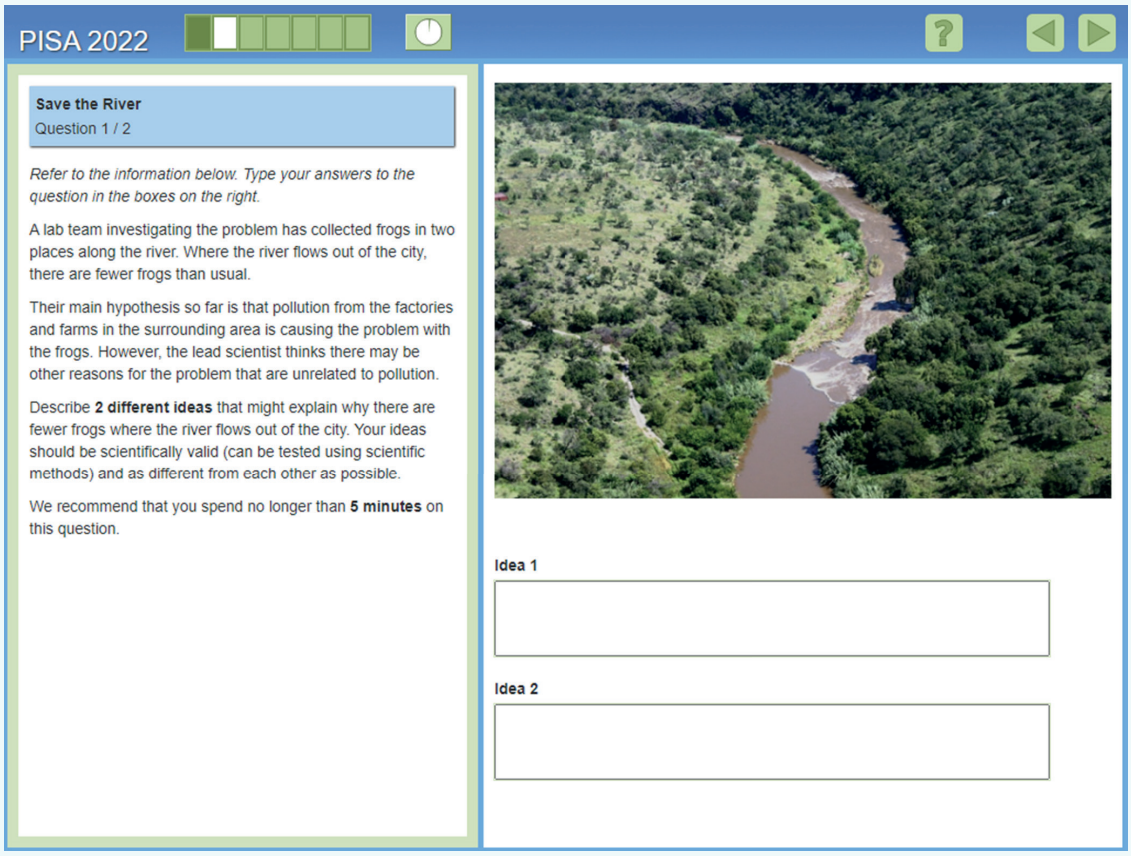
Task 6 – Save the River

- Ideation process: Generate diverse ideas
- Domain context: Scientific problem solving

The unit *Save the River* consisted of two tasks, and in this first task, students were asked to provide two different testable ideas that might explain a declining frog population in one part of a river flowing out of a city (Figure G). Students were explicitly asked to think of causes other than pollution from local farms and factories to encourage them to think of alternative hypotheses. As such, any ideas related to pollution were considered to have low appropriateness for this task.

Unlike the other divergent thinking tasks included in the PISA CT Rescoring project that ask students to propose three different ideas, students are only asked to provide two different ideas in this task.

Figure G. Save the River task



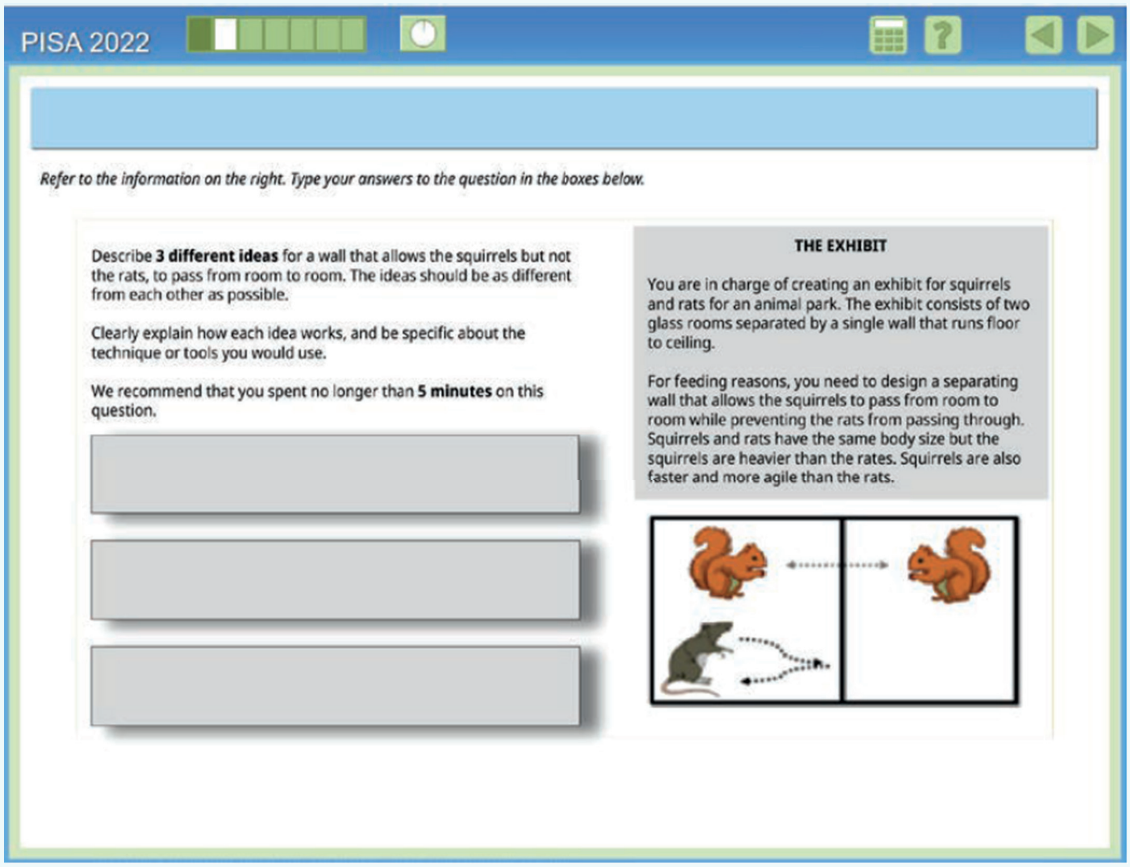
Note: This task is the first of two tasks in the *Save the River* unit.
Source: https://www.oecd.org/en/about/programmes/pisa/pisa-2022-creative-thinking-test-questions.html

Task 7 – The Exhibit

- Ideation process: Generate diverse ideas
- Domain context: Scientific problem solving

In *The Exhibit* task, students were asked to propose design ideas for constructing a separating wall in an animal park exhibit that would allow squirrels to pass from one room to the other while preventing the rats from doing so (Figure H). The item specified that while the squirrels and the rats had the same body size, squirrels are heavier, faster and more agile than the rats.

Figure H. The Exhibit task



Note: This task was the only task in the *The Exhibit* unit.
Source: PISA 2022 Creative Thinking test.

Note

1. In Belgium (French community) and Hong Kong (China, *English speaking*), minimum sample thresholds were not met due to the smaller sample size in that country-language group. In these cases, the maximum number of students was sampled. Given the very small sample size of the Hong Kong (China, *English speaking*) dataset, these data have been merged with the Hong Kong (China, *Chinese speaking*) dataset for most of the analyses conducted in this report. For more information on the PISA CT Rescoring project samples, see Annex A1.

References

Agogué, M. et al. (2014), "The impact of type of examples on originality: Explaining fixation and stimulation effects", *The Journal of Creative Behavior,* Vol. 48/1, pp. 1-12. [7]

Foster, N. and M. Piacentini (2025), "Assessing creative thinking at scale and across culture", in Katz-Buonincontro, J. and T. Kettler (eds.), *Oxford Handbook of Creativity and Education,* Oxford Handbooks, Oxford University Press. [3]

Foster, N. and M. Piacentini (forthcoming), "PISA 2022 creative thinking: What it is, what it is not, and what next?", *The Journal of Creative Behavior*, https://doi.org/10.1002/jocb.70068. [4]

Hatchuel, A. and B. Weil (2009), "CK design theory: An advanced formulation", *Research in Engineering Design*, Vol. 19, pp. 181-192. [6]

OECD (2025), *Creative minds in action: Students' imagination and ideas in storytelling, design and problem-solving tasks on the PISA test*, OECD Education Policy Perspectives, No. 120, OECD Publishing, https://doi.org/10.1787/5b718eb2-en. [5]

OECD (2024), *PISA 2022 Results (Volume III): Creative Minds, Creative Schools*, PISA, OECD Publishing, https://doi.org/10.1787/765ee8c2-en. [1]

OECD (2023), *PISA 2022 Creative Thinking Framework,* PISA, OECD Publishing, https://doi.org/10.1787/dfe0bf9c-en. [2]

01

What makes "creative ideas" creative?

Researchers in the field of creativity have long debated how to define creativity and how to evaluate the creative quality of ideas or products. In general, creative work is judged by the twin criteria of 1) novelty and 2) appropriateness and/or value (Runco and Jaeger, 2012[1]; Mumford, Lonergan and Scott, 2002[2]). While novelty refers to being unusual, uncommon, original, or outside established norms, appropriateness and value refer to qualities of relevance, usefulness, feasibility and efficiency. While there is relative consensus that each of these criteria are important, what is less clear is the extent to which they determine what counts as "creative". Research has found that the criteria tend to interact with each other when judges are asked to evaluate the overall creativity of an idea (Diedrich et al., 2015[3]) and that these interactions may differ across cultures and contexts (McCarthy, Chen and McNamee, 2018[4]).

From an educational perspective, understanding what makes ideas creative can help educators recognise creativity and creative thinking in different subject areas, as well as being able to teach students how to incubate creative ideas. For instance, some research findings have suggested that it is difficult for students to think both originally and appropriately at the same time (Runco, Illies and Eisenman, 2005[5]). Others have suggested that an effective way to manage this tension, as reported by professionals in creative industries, is to focus primarily on one dimension of creativity and allow the other to emerge subsequently (Harvey and Cronin, 2020[6]). Ivcevic and Kaufman (2024[7]) suggest that the meaning of the appropriateness and value criteria in defining creativity should be adapted to the level of creativity required for a given context, with meaningfulness and relevance being the most apt interpretation of these criteria in educational contexts whereas effectiveness and impact are most important in professional and eminent contexts. For contextualising originality in educational settings, Bahar and Maker (2025[8]) suggest adopting a standards-based approach: teachers can identify specific criteria linked to learning objectives in the curriculum to evaluate the originality of students' work, without needing to compare their students' work to a norm group or relying on expert consensus.

What do the PISA CT Rescoring project data tell us about what makes a creative idea creative? Is every creative idea appropriate? Is it more important for students to practise strategies for coming up with original ideas or is it better for them to focus on generating ideas that have value? Do these things change depending on the context?

1.1. Are creative ideas always appropriate?

Many have argued that appropriateness is a fundamental criteria for any idea to be creative. Logically speaking, it is reasonable to expect this to be the case: an idea that is highly original but is completely inappropriate does not serve much purpose. Yet research on the relationship between appropriateness and originality has produced mixed results, with some finding an inverse relationship between the two criteria (Runco, Illies and Eisenman, 2005[5]; Runco and Charles, 1993[9]) and others noting a dialectical tension between them that must be effectively balanced to produce creative outcomes (Rastelli et al., 2022[10]; Harvey and Cronin, 2020[6]). While these studies have focused on examining this relationship based on a relatively small sample size of adult or children subjects (all N<200) in specific contexts, we can investigate this question using a much larger and representative sub-sample of 15-year-olds using the PISA CT Rescoring project data (Box 1.1).

Box 1.1. Which score data are used in this chapter?

The analyses in this chapter draw upon two types of rescored data:

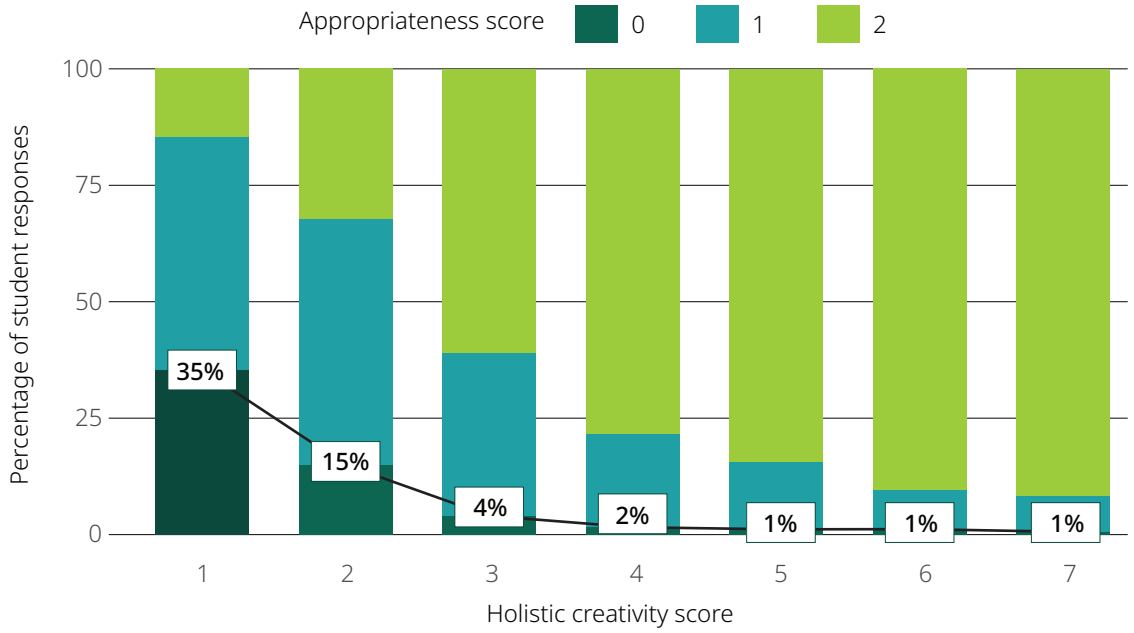
- Holistic creativity scores: Valid student responses are given a score of between 1 (lowest) and 7 (highest), based on judges' evaluations of the overall creative quality of the response. The response score reflects a relative ranking of creative quality compared to all other responses in that country-language-task group. Non-valid responses include those where all ideas are missing, aberrant or clearly inappropriate.
- Criteria scores: Each idea within a student response was scored for its appropriateness, originality and value. A cumulative idea sum score was also produced by aggregating the scores across these three criteria. Tasks classified as 'generate diverse ideas' tasks required students to submit multiple ideas per response.

The analyses in this chapter focus on the relative relationship between the two scoring methods – namely, which criteria drive increases in the holistic creativity scores. For more information on the scoring methods, see Annex A2.

Our findings show that, in general, as judges' holistic creativity scores increase so does the share of responses in that score category that are appropriate (Figure 1.1) – with this pattern exhibited more acutely at the lower end of the holistic score scale and plateauing towards the upper end. At the same time, appropriate answers do not automatically translate into high creativity scores: around 15% percent of responses that received a holistic creativity score of 1 (i.e. the lowest "valid" holistic score) were fully appropriate as were almost 80% of the responses in the mid-point of the scale (score of 4). Although almost no completely inappropriate responses were considered highly creative, defined in this chapter as scoring 5 or above in holistic score, a non-negligeable proportion of these were only partially appropriate. While these results may be due, in part, to the scoring and analytical methodology applied (see Box 1.1 and Box 1.2), it also suggests that deficiencies in idea appropriateness can in some circumstances be overcome by other factors (e.g. a high originality or value factor) and still be considered highly creative. It therefore seems to confirm from these initial findings that creative work is usually appropriate – at least, partially so – but that appropriateness alone is insufficient for producing creative outcomes.

Figure 1.1. High-scoring holistic creativity responses tend to be highly appropriate

Share of responses with each appropriateness score amongst all responses (across tasks and country-language groups), per holistic score category



Note: Percentages were computed on the pooled responses across all tasks within each country-language group and then averaged across country-language groups. The percentages shown in the Figure labels refer to the share of responses in each holistic score category that scored 0 points in appropriateness. For responses to items classified as 'generate diverse ideas' items that require students to submit multiple ideas, the appropriateness scores for all ideas within the response were first averaged to give a single appropriateness score per student response. The averaged appropriateness scores were then rounded to the nearest integer (0, 1 or 2).

Source: Table B1.1

Box 1.2. Caution on interpreting the data in this chapter

Results reported in this chapter should be interpreted bearing in mind two points: (1) the principles of the two scoring methods involved; and (2) the analytical approach used in the chapter (i.e. averaging idea criteria scores in divergent thinking tasks).

Principles of the holistic and criteria-based scoring methods: relative ranking versus absolute scores

The two scoring methods compared in this chapter differ in how judges assign scores. The holistic judgement method, from which the holistic creativity scores are derived, involves ranking student responses in each country-language-task group *relative to other responses* within the same country-language-task group. In contrast, the criteria-based method assigns scores to student ideas independently of the quality of other ideas in the sample. These differences between the two scoring approaches may mean that, in country-language groups that have relatively more high performing students in general, it may be more likely that fully appropriate responses are assigned a low holistic creativity score than in jurisdictions where students struggle to think creatively. In other words, if most ideas in the country-language-task sample are appropriate, then some low creativity responses will likely be appropriate. Likewise, in country-language groups where students perform more poorly overall compared to other jurisdictions, it may be more likely that high holistic creativity responses are only partially appropriate. This feature should be kept in mind when directly comparing the relative proportions of appropriate responses per holistic score category across country-language groups. The same caution applies to direct comparisons involving the other score criteria (originality, value) described in this chapter.

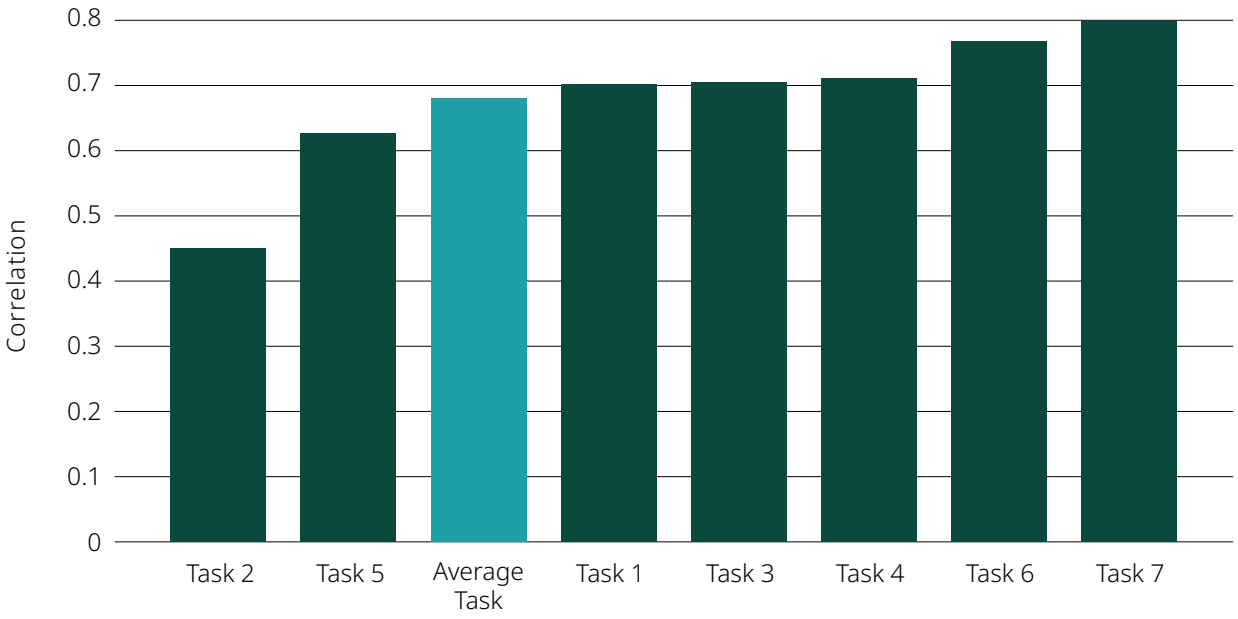
Analytical approach: averaging criteria-based scores across ideas in divergent thinking tasks

The analytical approach employed in this chapter assumes that each idea within a student response in divergent thinking tasks contributes equally to a judge's holistic creativity score. The criteria scores are therefore averaged across all ideas in a response before then being associated to the response holistic creativity scores. However, it may be that some ideas are more important than others for determining judges' overall holistic response score.

The role of appropriateness in influencing whether responses are creative also seems to vary across tasks (Figure 1.2). Although there is a strong association (r=0.68) between appropriateness and holistic creativity scores, on average across country-language-task groups, this correlation is only moderate (r=0.45) for Task 2 – *Illustration Titles* (see Insert in the Introduction of this report for more information on the tasks included in the PISA CT Rescoring project). In some country-language groups, there is even no significant association between the two scores for Task 2. In general, the problem-solving tasks (Tasks 4-7, Figures E to H in the Introduction) tend to show a greater correlation between holistic creativity and appropriateness scores than the expressive tasks in the study (Tasks 1-3, Figures B to D in the Introduction).

Figure 1.2. Holistic creativity scores and appropriateness scores are strongly correlated

Average correlation across country-language groups, by task



Note: All correlations are statistically significant at the average country-language level. The average task results were computed by task and then averaged across the 7 tasks. For responses to items classified as 'generate diverse ideas' items, the appropriateness scores for all ideas in a response were first averaged to give a single appropriateness score per student response.

Source: Table B1.2

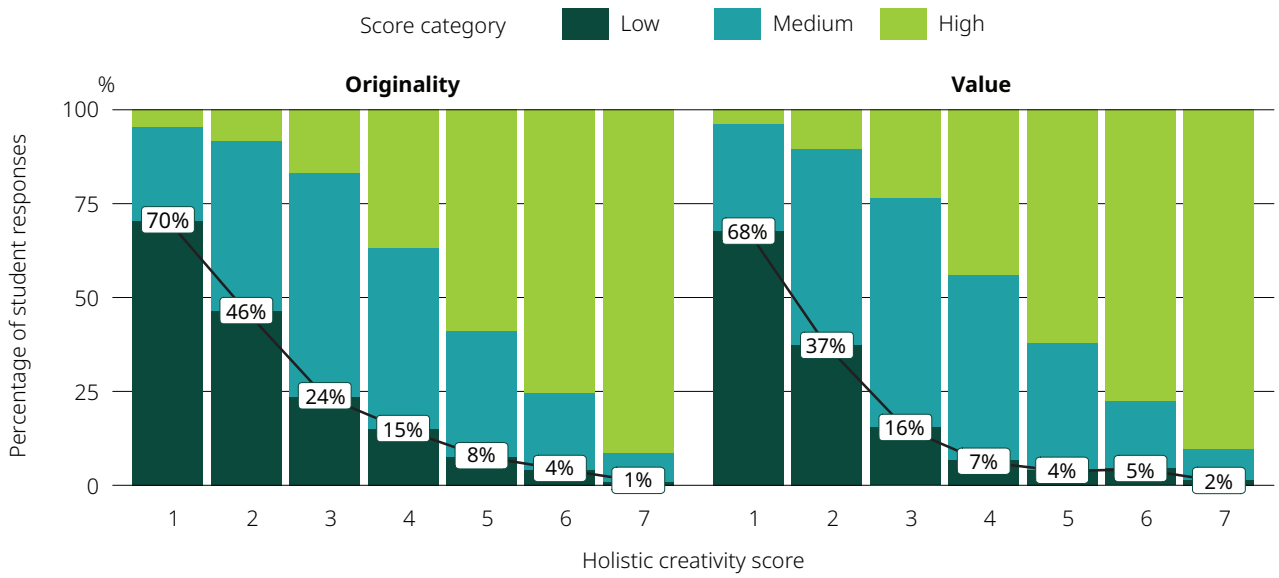
Why might these differences across tasks exist? Certain tasks may elicit more abstract ideas whose appropriateness are evaluated more subjectively by judges. For example, in the *Illustration Titles* task, students were asked to suggest three different title ideas for a fantasy countryside scene (see Figure C in Introduction). In this context, relatively abstract idea associations may be perceived as creative even if not wholly appropriate given the task places a premium on imagination. It is also reasonable that what constitutes an appropriate answer for this task – that asks students to make idea associations to elements of an already abstract and whimsical image – is more a matter of subjective interpretation than in other more objective task contexts (e.g. problem-solving tasks). That is not to say that all creative expression tasks have low appropriateness requirements simply because they require imagination: in fact, both the remaining written and visual expression tasks included in study, respectively, demonstrated a strong correlation between appropriateness scores and holistic scores ($r > 0.7$). Unlike the *Illustration Titles* tasks, both *2983* (Figure D) and *Science Fair Poster* (Figure B) tasks have well-defined constraints that might lead to a stricter interpretation of appropriateness: for example, in *2983*, students must suggest a story idea that concretely connects to the number 2983, and in *Science Fair Poster,* students must design a poster for a school science fair focused on a specific theme ('Life in Deep Space').

1.2. Originality and value: which is more important for determining creative ideas?

In addition to being appropriate, creative ideas tend to be original and valuable. But which factor is more important, and does this change across tasks? When comparing the distribution of original and valuable responses in each holistic score category, respectively, the results are very similar (Figure 1.3). Broadly speaking, as judges' evaluations of response originality and value increase, so do their overall holistic creativity scores. However, this trend is far less pronounced for responses with a relatively low holistic creativity score (e.g. scores 1 to 3). Contrast this pattern with the one observed in Figure 1.1: almost 61% of responses with a holistic creativity score of 3 are highly appropriate, compared to just 17% of responses in the same holistic score category that are highly original (Table B1.1 and Table B1.3). Conversely, the relative proportions of highly original or highly valuable responses increase dramatically as the holistic score increases from scores 4 to 7.

Figure 1.3. High-scoring holistic creativity responses also tend to have high originality and value

Share of responses with low, medium and high original/value scores amongst all responses (across tasks and country-language groups), per holistic score category



Note: Percentages were computed on the pooled responses across all tasks within each country-language group and then averaged across country-language groups. The percentages shown in the Figure labels refer to the share of responses in each holistic score category that were low original or value responses, respectively. 'Low' originality/value responses are those with an average idea criteria score of less than 1 point, 'medium' originality/value responses are those with an average idea criteria score of at least 1 point but less than 2 points, and 'high' originality/value responses are those with an average idea criteria score of 2 or more points. For 'generate diverse ideas' items that contain multiple ideas per response, the originality/value scores for all ideas were first averaged to give a single criteria score per student response. All originality/value scores for non-appropriate ideas were recoded to 0 for these analyses.

Source: Table B1.3 and Table B1.4

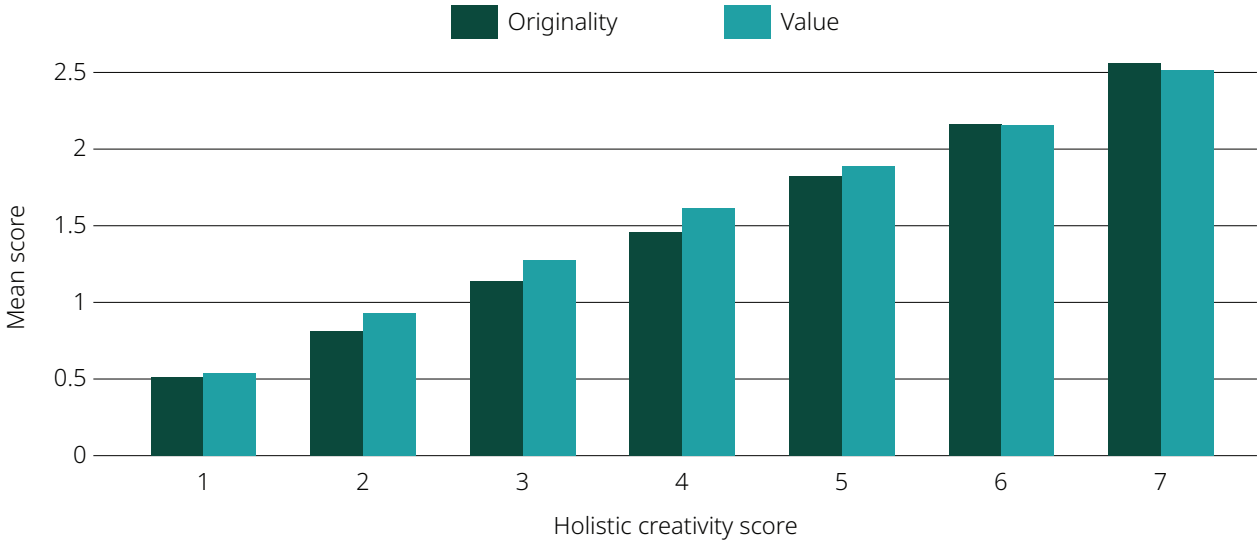
How should these results be interpreted? While appropriateness might act as an "entry barrier" for coming up with a creative idea, once this criterion is partially achieved it seems to add relatively little to a response's holistic creativity – and indeed, many non-creative ideas can be highly appropriate. In contrast, relatively few low creativity ideas are either highly original or highly valuable. For the minority that are, their relatively low holistic creativity score can likely be explained by the absence of the other factor. For example, amongst highly original responses, those that overall have a low holistic creativity score record significantly lower value scores on average across country-language groups (mean difference of 0.7 score points) (Table B1.5). While the same is true for appropriateness, the differential is much smaller (close to 0.2 score points). A similar mean difference in originality scores is observed amongst high value responses that have a low holistic creativity score (mean gap of 0.7 score points), with an even smaller gap to the mean appropriateness scores of those responses (difference of 0.1 score points). While the latter is still significant on average across country-language groups, this only holds in a third of the individual country-language groups (Table B1.6).

Despite their similarities, can we identify differences in the patterns of original and valuable responses at different holistic creativity scores? When highly and moderately original responses are considered together, the increase in their relative proportion within each holistic creativity score increases quite consistently as the holistic scores increase (see the slope of the line in Figure 1.3). When looking at the same pattern amongst at least moderately valuable responses, the relationship is slightly less linear: while the share of at least somewhat valuable responses increases at a larger rate amongst low holistic creativity responses (an increase of around 50 percentage points from holistic score category 1 to 3), the proportion plateaus around the mid-range with over 93% of all responses at holistic score 4 or above considered at least somewhat valuable, and increasing by just 5 percentage points to 98% at holistic score 7 (Table B1.3). In comparison, only 85% of responses at holistic score 4 are at least moderately original, but this rises by 14 percentage points to 99% of all responses awarded the top holistic score of 7 (Table B1.4).

Overall, the PISA CT Rescoring study findings seem to confirm that both originality and value criteria are important for defining whether an idea is creative but with two caveats: 1) that the holistic creativity scores are more sensitive to response value at the lower end of the score spectrum than at the upper end; and 2) that it is more likely that the most creative responses are highly original than they are highly valuable. These findings are also implied when looking at the mean originality and value scores of responses in each holistic category: response value scores are greater amongst those with holistic creativity scores 5 or below, but response originality scores are greater amongst those with holistic creativity scores of 6 or 7 (Figure 1.4). This result might also be explained by the relative difficulty of achieving each criteria – in other words, it might be easier to suggest an effective idea than an original one – but it also suggests that originality matters slightly more for distinguishing somewhat creative ideas from very creative ones.

Figure 1.4. The most creative answers are more original than they are valuable

Mean response originality and value scores (across tasks and country-language groups), by holistic score category



Note: Mean original and value scores were computed on the pooled responses across all tasks within each country-language group and then averaged across country-language groups. For 'generate diverse ideas' items that contain multiple ideas per response, the originality/value scores for all ideas were averaged to give a single criteria score per student response. All originality/value scores for non-appropriate ideas were coded as 0.

Source: Table B1.7 and Table B1.8

1.2.1. Differences across tasks

Are originality and value similarly associated with creative outcomes in different types of task? As expected, both response originality and value scores are significantly and strongly correlated with response holistic creativity scores on average across tasks and country-language groups (r = 0.73 for originality and r = 0.72 for value; Table B1.9 and Table B1.10). Within country-language groups, the two criteria tend to demonstrate an equally strong association with the holistic creativity scores – though there are some exceptions. For example, in Canada (French provinces), originality is more strongly correlated with holistic creativity scores (difference of 0.2 compared to the equivalent correlation for value) but in Greece the opposite is observed.

Across the seven different tasks included in the PISA CT Rescoring project, on average across country-language groups, these relationships are significant and moderate to strong (range of 0.65 to 0.79). The task that records the weakest correlation between holistic score and originality score is *Illustration Titles* (a written expression task; see Figure C in the Introduction) (r=0.66) and the task that records the strongest correlation between the two scores is *The Exhibit* (an engineering problem-solving task; see Figure H in the Introduction) (r=0.78). At first glance, these results might seem unexpected: it would be reasonable to expect originality to be more influential in determining creativity scores in more imaginative task contexts and less so in narrower problem-solving tasks; yet it may be precisely for that reason that original solutions in *The Exhibit* task are perceived as more creative. It is relatively easier to come up an uncommon or imaginative

idea association in an open, abstract task compared to one that requires a functional solution. That is not to say that originality is not important in expressive tasks, but rather, it may be more common for student responses to be imaginative, and therefore originality is less of a feature that separates poor ideas from average or highly creative ideas. We can infer the relative difficulty of the tasks with respect to each criteria by looking at the mean criteria-based scores per task (Tables B1.11-B1.17). The mean sum score and originality score for *Illustration Titles* is significantly higher than for all other tasks included in the study, on average across countries, with the opposite being true for *The Exhibit*. For example, students score over 1.5 points more in total and over 0.5 points more in originality in the former, than they do in the latter (Table B1.12 and Table B1.17).

A similar effect of task difficulty could explain the relationship between value and overall holistic creativity scores. In our study, response value scores were least correlated with holistic creativity scores in the *Food Waste* (social problem-solving) task (r=0.6) and most strongly correlated in *The Exhibit* task (r=0.79). Again, it may be intuitive to expect value would be most strongly correlated with problem-solving tasks regardless of task difficulty, but this is not what we observe. The *Food Waste* item asks students to think about an everyday problem that is likely to be a familiar context and one for which students may already have some awareness or lived experience; as such, students may find it easier to suggest ideas that would be functional and effective. In fact, out of the four problem-solving items included in the PISA CT Rescoring project, students are most successful overall, and in coming up with valuable ideas, in the *Food Waste* task (Figure F in the Introduction) (Tables B1.14-B1.17). In contrast, *The Exhibit* task is less familiar and potentially more demanding from a knowledge perspective (e.g. knowledge of basic physics, engineering or biological principles); as a result, we observe that it is the problem-solving task in which students are least successful overall.

1.3. Creative ideas: greater than the sum of their parts?

While most researchers agree that appropriateness, originality and value are all fundamental characteristics of creative ideas, these criteria alone do not fully explain what makes creative ideas creative. When considered together, these criteria scores only explain around two-thirds (66%) of the total variance in holistic creativity scores, on average across country-language-task groups (Table B1.18). These findings align with previous research that has demonstrated that creativity scores are influenced by more factors than originality and quality (Long, 2014[11]), such that the creative value of the whole might be considered greater than the sum of its parts. While criteria-based (sometimes referred to as analytic) methods can provide a more easily understood and replicable formula for evaluating creative ideas in a measurement context, they may fail to capture the full creative quality of ideas in the way that more holistic scoring methods are able to.

There are many other factors – individual-, task- and judge-related – that might determine how creative ideas are formed, perceived and evaluated. For instance, domain-specific theories of creativity assert that domain-relevant knowledge and skills are important determinants of successful creative thinking processes. While domain readiness may certainly help individuals to come up with ideas of value, and perhaps even original ideas, these scoring criteria may fail to fully capture the role of relevant knowledge and skills in supporting creative thinking. Runco, Illies and Eisenman (2005[5]) also demonstrated how task instructions and context can differentially influence individuals' success in generating original and flexible ideas, and the analysis in this chapter also suggests that task difficulty plays a mediating role in how judges differentially perceive the importance of the same criteria in different tasks – even within the same broader domain context. Several studies have also demonstrated that rater expertise (Kaufman et al., 2008[12]; Kaufman, Baer and Cole, 2009[13]; Long, 2014[11]), subjective and socio-cultural rater bias (Lloyd-Cox, Pickering and Bhattcharya, 2022[14]; Kaufman et al., 2010[15]), and ambiguity in how the scoring criteria are defined for given tasks (Reiter-Palmon et al., 2009[16]) can influence how judges understand and evaluate creative ideas. Assessment choices, including how to operationalise scoring approaches, should be mindful of these factors and take into account the nature of the assessment tasks and the dimension(s) of interest of the construct (Reiter-Palmon, Forthmann and Barbot, 2019[17]).

Key findings and implications

- Creative ideas are not always highly appropriate, highly original and highly valuable – but they very likely are.
  - Over 88% of highly creative ideas are appropriate, and over 75% are highly original and highly valuable.
- The PISA CT Rescoring project shows that it is possible to evaluate differences in student idea quality across the criteria of appropriateness, originality and value in a variety of everyday imaginative and problem-solving tasks.
  - Expectations of what appropriateness, originality and value mean in educational contexts should be set at developmentally appropriate levels and should focus on identifying evidence of "little-c" creativity rather than the expression of creative genius. For example, focusing on relevance and meaningfulness in students' ideas might be the most apt interpretation of the appropriateness and value criteria (Ivcevic and Kaufman, 2024[7]). For originality, focusing on uncommon ways that students demonstrate their understanding of specific learning objectives grounded in curricular expectations may be the most instructionally relevant approach (Bahar and Maker, 2025[8]).
- The relative importance of appropriateness, originality and value varies across task types, though they remain important in all contexts.
  - Task demands, constraints and difficulty appear to affect how much each criteria determines the overall creativity of a response.
- Appropriateness, originality and value do not fully explain what makes creative ideas creative, accounting for around two-thirds of the variation in holistic creativity scores on average. Other factors also play a role in determining the extent to which individuals successfully achieve creative outcomes across different tasks. These factors might include:
  - Domain-relevant knowledge and skills that support students' contextualisation of appropriate, original and valuable ideas;
  - Students' background, characteristics, or attitudes and traits that might support or hinder the development of creative thinking skills. Some of these are explored further in subsequent chapters in this report (e.g. attitudes and traits in Chapter 3; gender in Chapter 6; and socio-economic status in Chapter 7).

References

Bahar, K. and C. Maker (2025), "A new perspective on scoring children's originality: a standards-based criterion-referenced assessment approach", *Frontiers in Psychology*, Vol. 16, https://doi.org/10.3389/fpsyg.2025.1545396. [8]

Diedrich, J. et al. (2015), "Are creative ideas novel and useful?", *Psychology of Aesthetics, Creativity, and the Arts*, Vol. 9/1, pp. 35-40, https://doi.org/10.1037/a0038688. [3]

Harvey, S. and M. Cronin (2020), *Make it work: The importance of usefulness and appropriateness for creativity,* https://doi.org/10.5465/AMBPP.2020.20251abstract. [6]

Ivcevic, Z. and J. Kaufman (2024), "Inside the box: Considering the role of appropriateness for creativity", *Psychology of Aesthetics, Creativity, and the Arts*, https://doi.org/10.1037/aca0000703. [7]

Kaufman, J., J. Baer and J. Cole (2009), "Expertise, domains, and the consensual assessment technique", *Journal of Creative Behavior*, Vol. 43/4, pp. 223-233, https://doi.org/10.1002/j.2162-6057.2009.tb01316.x. [13]

Kaufman, J. et al. (2008), "A comparison of expert and nonexpert raters using the consensual assessment technique", *Creativity Research Journal*, Vol. 20/2, pp. 171-178, https://doi.org/10.1080/10400410802059929. [12]

Kaufman, J. et al. (2010), "In the eye of the beholder: Differences across ethnicity and gender in evaluating creative work", *Journal of Applied Social Psychology,* Vol. 40/2, pp. 496-511, https://doi.org/10.1111/j.1559-1816.2009.00584.x. [15]

Lloyd-Cox, J., A. Pickering and J. Bhattcharya (2022), "Evaluating creativity: How idea context and rater personality affect considerations of novelty and usefulness", *Creativity Research Journal*, Vol. 34/4, pp. 373-390. [14]

Long, H. (2014), "More than appropriateness and novelty: Judges' criteria of assessing creative products in science tasks", *Thinking Skills and Creativity,* Vol. 13, pp. 183-194, https://doi.org/10.1016/j.tsc.2014.05.002. [11]

McCarthy, M., C. Chen and R. McNamee (2018), "Novelty and usefulness trade-off: Cultural cognitive differences and creative idea evaluation", *Journal of Cross-Cultural Psychology,* Vol. 49/2, pp. 171-198, https://doi.org/10.1177/0022022116680479. [4]

Mumford, M., D. Lonergan and G. Scott (2002), "Evaluating creative ideas: Processes, standards and context", *Inquiry: Critical Thinking Across Disciplines,* Vol. 22/1, pp. 21-30, https://doi.org/10.5840/inquiryctnews20022213. [2]

Rastelli, C. et al. (2022), "Balancing novelty and appropriateness leads to creative associations in children.", *PNAS Nexus,* Vol. 1/5, https://doi.org/10.1093/pnasnexus/pgac273. [10]

Reiter-Palmon, R., B. Forthmann and B. Barbot (2019), "Scoring divergent thinking tests: A review and systematic framework", *The Psychology of Aesthetics, Creativity, and the Arts,* Vol. 13/2, pp. 144-152, https://doi.org/10.1037/aca0000227. [17]

Reiter-Palmon, R. et al. (2009), "Creativity and domain specificity: The effect of task type on multiple indexes of creative problem-solving.", *Psychology of Aesthetics, Creativity, and the Arts*, Vol. 3/2, pp. 73-80, https://doi.org/10.1037/a0013410. [16]

Runco, M. and R. Charles (1993), "Judgements of originality and appropriateness as predictors of creativity", *Personality and Individual Differences,* Vol. 15/4, pp. 537-546, https://doi.org/10.1016/0191-8869(93)90337-3. [9]

Runco, M., J. Illies and R. Eisenman (2005), "Creativity, originality and appropriateness: What do explicit instructions tell us about their relationships?", *Journal of Creative Behavior,* Vol. 39/2, https://doi.org/10.1002/j.2162-6057.2005.tb01255.x. [5]

Runco, M. and G. Jaeger (2012), "The standard definition of creativity", *Creativity Research Journal*, Vol. 24/1, pp. 92-96, https://doi.org/10.1080/10400419.2012.650092. [1]


x4

02

Divergent and convergent thinking: which tasks led to more creative ideas on the PISA test?

The literature broadly agrees that creative thinking involves both divergent-exploratory and convergent-integrative cognitive processes (Barbot, Besançon and Lubart, 2011[1]; Cropley, 2006[2]; Dygert and Jarosz, 2020[3]; Guilford, 1967[4]). Divergent thinking supports idea generation by enabling individuals to connect diverse pieces of information and adopt different points of view (Guilford, 1956[5]; Guilford, 1967[4]). Convergent thinking is the ability to apply logical reasoning to information and it is crucial for problem-finding, identifying and selecting good ideas, and supporting decision-making (Guilford, 1956[5]; Cropley, 2006[2]). This process has also been referred to as convergent-integrative thinking, where several disparate elements are brought together into a new, original synthesis (Barbot, Besançon and Lubart, 2011[1]). Some researchers have also sought to understand and explore the similarities between the dual cognitive processes involved in creative thinking and other dual-process models of cognition that differentiate between more freely associative and more structured or deliberate thinking modes (e.g. (Sowden, Pringle and Gabora, 2018[6]; Smith, Ward and Finke, 1995[7]; Beaty et al., 2016[8]). Generally, these dual divergent and convergent cognitive processes involved in creative work occur iteratively throughout the creative process, but their relative importance and effectiveness may vary depending on the domain or task at hand (Baer, 2016[9]; Barbot, Besançon and Lubart, 2011[1]; Botella and Lubart, 2015[10]), situational or dispositional factors (Nijstad et al., 2010[11]), or the stage of the creative process in which they occur (Howard-Jones, 2002[12]).

The PISA 2022 test identified and measured three distinct ideation processes that reflected these divergent-exploratory and convergent-integrative aspects of creative thinking, with each test task corresponding to a specific ideation process (OECD, 2023[13]). 'Generate diverse ideas' tasks elicited evidence of the divergent-exploratory aspects of creative thinking: these tasks presented students with a context and some stimulus material (e.g. an image, a problem or need, etc.), and asked students to come up with two or three ideas that were as different as possible from each other. 'Generate creative ideas' and 'evaluate and improve ideas' tasks similarly presented students with a context and stimulus material but asked them to come up with an original idea or improve someone else's idea. While these latter task types also involve idea generation, they also involve the convergent-integrative aspects of creative thinking (understanding the problem, evaluating and selecting ideas, and/or producing an original synthesis).

The PISA 2022 results reported students' success in the different types of task in the assessment (i.e. in generating diverse or original ideas), but the results could not provide information about the overall *creative quality* of students' ideas resulting from these different ideation processes. Data from the PISA CT Rescoring project does allow us to examine this. Can we observe differences in the quality of creative outcomes following different ideation processes? What might any differences imply for supporting students to come up with creative ideas in different types of tasks?

2.1. Do divergent thinking tasks generate creative ideas?

Four of the seven tasks included in the PISA CT Rescoring project were classified as 'generate diverse ideas' tasks: *Illustration Titles*, *Food Waste* and *The Exhibit* (Tasks 2, 5 and 7) asked students to come up with three ideas for a given scenario that were as different as possible, whereas the *Save the River* task (Task 6) asked students to come up with just two different ideas (see Figures C, F, G and H in the Introduction section of this report for more information on the tasks included in the study). Although the task instructions did not explicitly instruct students to think of creative ideas *per se* in these tasks, did students generate them anyway as a result of engaging in divergent thinking?

Figure 2.1 shows the share of students that suggested none, one or multiple creative ideas in the four 'generate diverse ideas' tasks included in the PISA CT Rescoring study. "Creative ideas" in this chapter are defined as those that scored highly (within 1 point of the maximum score) for each idea criteria of appropriateness, originality and value (see Box 2.1). On average across country-language groups, just under half (48%) of all student responses in 'generate diverse ideas' tasks requiring three ideas included at least one creative idea, despite students not being explicitly instructed to be creative in their responses. Around 1 in 4 student responses to these divergent thinking tasks included at least two creative ideas, on average across country-language groups (Table B2.1).

Box 2.1. Which score data are used in this chapter?

The analyses in this chapter draw upon one type of rescored data from the PISA CT Rescoring project, the **criteria scores**. The Criteria-Based scoring method awarded each idea within a student response a score for appropriateness, originality and value (see Annex A2). A cumulative idea sum score was also produced by aggregating the scores across these three criteria for each idea.

This chapter focuses on examining the quality of students' ideas in subsets of tasks ('generate diverse ideas' and 'generate creative ideas' tasks) that involve different ideation processes. The analyses conducted in this chapter identify a "**creative idea**" as an idea that scored at least 1 point for appropriateness (out of an available 2 score points) and at least 2 points for both originality and value respectively (out of an available 3 score points each).

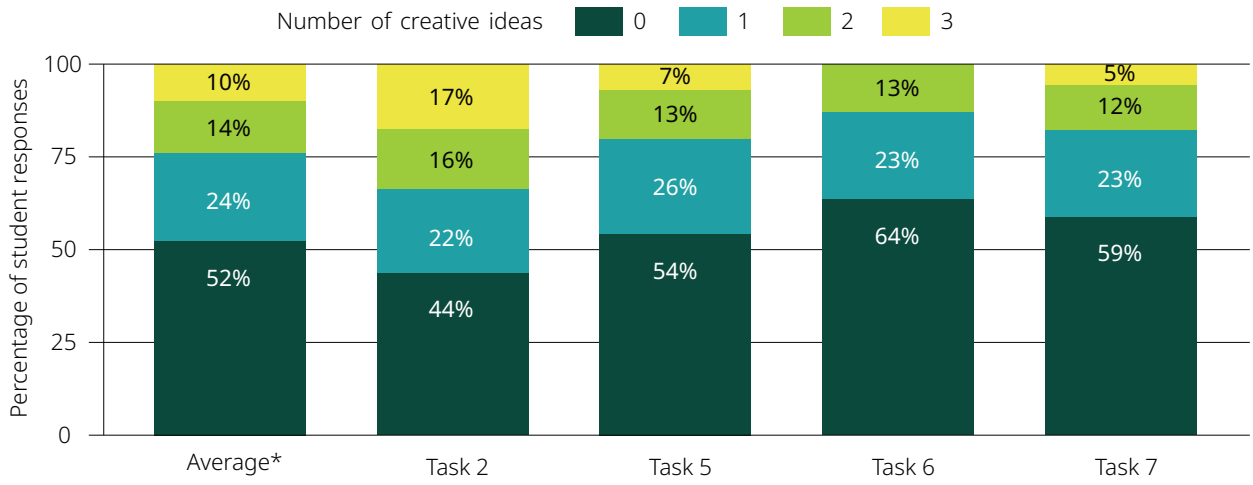
We use criteria scores for the analyses in this chapter as these scores apply to each idea within a student response. In contrast, only one holistic creativity score is attributed per student response (i.e. considering all ideas together), and as such, these scores do not allow the quality of each idea within a response to be examined independently.

In general, and in around half of the country-language groups, students were less likely to propose creative ideas in the problem-solving tasks asking for different ideas (Tasks 5, 6 and 7) compared to the sole written expression divergent thinking task, *Illustration Titles* (Task 2) (Table B2.1). This difference may largely be due to the relative difficulty of the tasks in the different domains, with students more likely to achieve lower sum scores and individual criteria scores in the problem-solving tasks on average across country-language groups than in the creative expression tasks (Tables B2.2).

Across country-language groups, there is significant variation in the number of responses containing creative ideas in the 'generate diverse ideas' tasks (Figure 2.2). Students in Korea were by far the most successful, with nearly 9 in 10 student responses including at least one creative idea on average across the three tasks requiring three ideas (Tasks 2, 5 and 7). Well over 50% of student responses in Hong Kong (China), Italy, Saudi Arabia, Slovenia, Belgium (French community), Germany and France also contained at least one creative idea on average across these three tasks. Conversely, over 80% of responses in Colombia and Brazil, and nearly 90% of responses in Chile, did not include any creative ideas on average across the three 'generate diverse ideas' tasks requiring three ideas.

Figure 2.1. Nearly half of all student responses in 'generate diverse ideas' tasks contained at least one creative idea

Percentage of student responses with 0-3 creative ideas, on average across country-language groups, by task



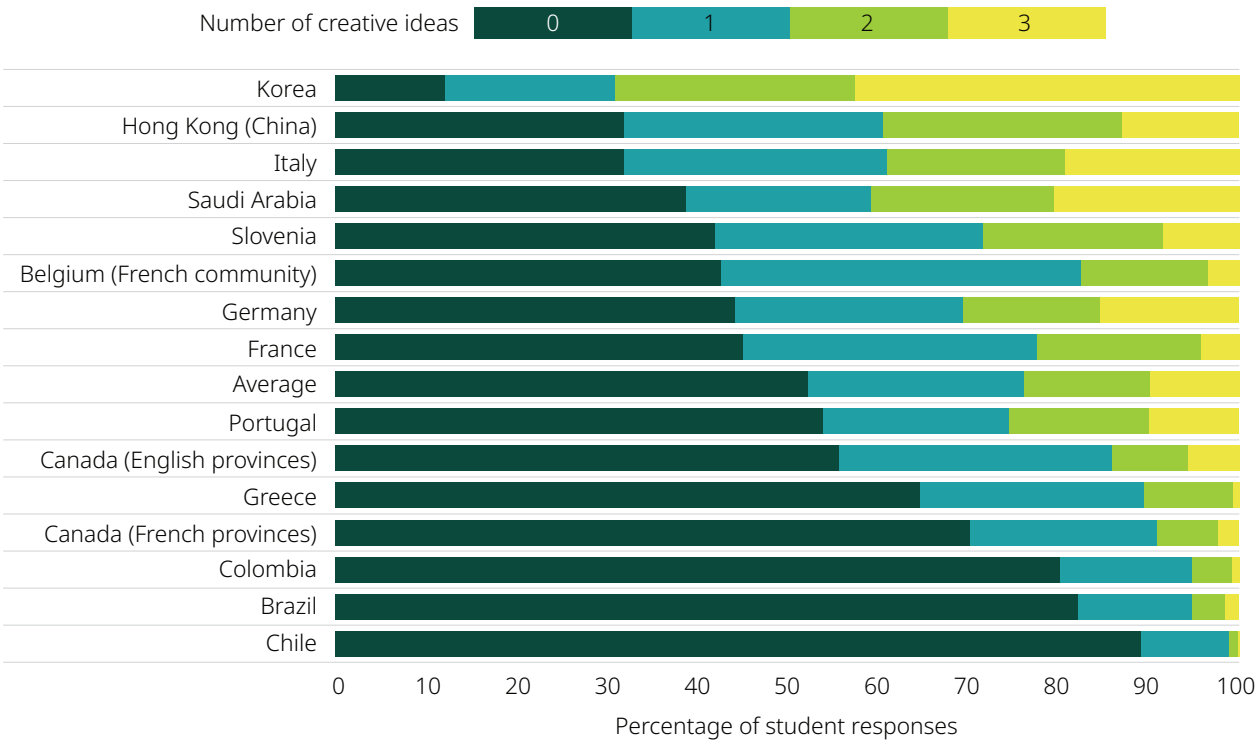
Note: (*) As Task 6 - *Save the River* requires only two ideas, rather than three ideas, the Average task bar excludes data from Task 6. Average percentages for each task were computed by country-language group then averaged across country-language groups. 'Creative ideas' are defined as those that scored at least 1 score point in appropriateness (out of a possible 2 score points) and that scored at least 2 score points in both originality and value (out of a possible 3 score points each).

Source: Table B2.1

These results suggest that even when students are not explicitly prompted to think of *creative* ideas, specifically, divergent thinking tasks nonetheless encourage creative idea generation. These findings also align with experimental studies focused on training creativity that have generally targeted divergent thinking processes and noted positive effects (see Box 2.2). It is notable that many students across many country-language group contexts in our study can successfully think of creative ideas in these tasks even though they only require a limited number of different ideas (in contrast to, for example, the "think-of-as-many-ideas-as-you-can" model that is typically used in other divergent thinking tasks). It may be that students engaged in some initial filtering of a longer list of ideas before submitting their responses to these PISA 2022 tasks, thus selecting only their 'best' ideas, but it suggests that even limited idea generation activities can support creative outcomes.

Figure 2.2. The share of student responses containing creative ideas varies across country-language groups

Percentage of student responses with 0-3 creative ideas, on average across 'generate diverse ideas' tasks (*), by country-language group



Note: (*) As Task 6 - *Save the River* requires only two ideas, rather than three ideas, the country-language group averages exclude data from Task 6. 'Creative ideas' are defined as those that scored at least 1 score point in appropriateness (out of a possible 2 score points) and that scored at least 2 score points in both originality and value (out of a possible 3 score points each).

Source: Table B2.1

Box 2.2. Research spotlight: Strategies for teaching students to think creatively

Various creativity techniques have been developed to enhance creative thinking at different stages of the creative process, although most focus on facilitating idea generation (Vernon, Hocking and Tyler, 2016[14]). In general, techniques aiming to promote idea generation generally employ two types of methods: (1) a stimuli-oriented method (i.e. using input to achieve a shift in perspective and foster creative idea generation) and (2) a relationship-oriented method (i.e. focusing on free association or forced relationships as means to generate new ideas). For instance, studies have found a positive effect of showing novel examples (in contrast to common examples) as input stimuli, in particular to improve the originality of ideas generated during divergent thinking exercises (Agogué et al., 2014[15]; Yuan et al., 2021[16]), although others have highlighted that input stimuli should be aligned with individual needs and preferences in order to be effective (De Jonge, Rietzschel and Van Yperen, 2018[17]).

Relatively less research has focused on techniques to promote and improve convergent thinking processes. Some techniques, such as idea evaluation metrics or strengths and weaknesses analyses, have been used in prior creativity training programs but were found to have little effect on convergent thinking outcomes – likely because success in identifying and evaluating "good" ideas is intimately tied with domain knowledge, while creativity techniques tend to address creative thinking in a domain-general way (van Broekhoven et al., 2020[18]; Ritter et al., 2020[19]).

General creativity training interventions in higher education settings

Most recent studies examining the effectiveness of creativity training programs have been conducted in university settings, employing general cognitive-based techniques to promoting students' creativity. Ritter et al. (2020[19]) conducted a year-long creativity training program involving university students in the Netherlands enrolled in the Business department, during which students learned how to apply the Six Step Cycle of Creativity (understanding the question, convergent thinking, divergent thinking, detached thinking, stop thinking, and sleeping) to a range of problems. The study found the intervention significantly improved students' idea generation and flexibility in divergent thinking tasks, but not idea originality. Interestingly, although the course ran for an academic year, researchers observed significant effects after just 3 months of training, with no further improvement in students' creative performance following further training. In a separate study, Ritter and Mostert (2017[20]) also found that even a brief 1.5 hour cognitive-based training improved outcomes on different creative thinking measures, particularly improving cognitive flexibility on divergent thinking tasks.

Similar undergraduate creativity and innovation training modules for students across disciplines have been developed by researchers at universities in the United Arab Emirates (UAE) and Germany, respectively. In the UAE study, the module introduced students to theoretical conceptualisations of creative thinking as well as to the practical application of a variety of techniques to solve real-world problems, and sessions primarily centred around group activities; after the training, students significantly improved in their idea originality, elaboration, and fluency (Vally et al., 2019[21]). In the German study, students attended eight

90-minute lectures to introduce participants to essential theories and models of creativity, before completing a 4-week practical component; results showed a positive effect of the training on creative performance, with the increase in performance remaining relatively stable over time (beyond the training period) (Meinel et al., 2018[22]).

School-based interventions aimed at improving students' scientific creativity

Most studies looking at the effects of creativity interventions in school-age populations have focused on creativity in the sciences. For example, Sun, Wang and Wegerif (2020[23]) conducted an experimental study involving 105 high school students in China. Students in the experimental condition were trained in divergent thinking strategies, including association, decomposition and combination with adjustment, with examples of how to apply these strategies to real-world problems. The study found a significant difference in the post-test performance between the experimental and control groups, suggesting that the cognitive-based training was useful in improving students' scientific creativity. The study found that all students (both those with "low" and "high" self-reported creative potential) benefitted from the training, but that students with greater scientific-domain knowledge before the training improved their scientific creativity even more than students with a low level of domain-knowledge.

Yang and colleagues (2016[24]) studied the effect of a well-designed, inquiry-based teaching classroom intervention over the course of a semester on students' scientific creativity in an elementary school in Chinese Taipei. They found that the intervention group significantly outperformed students in the control group in the post-tests focused on science inquiry, convergent thinking and divergent thinking respectively. The study also explored effective teaching practices on creative cognitive processes. The main practices supporting divergent thinking were 'facilitating associative thinking' and 'sharing impressive ideas', while 'encouraging evidence-based conclusions' and 'reviewing and commenting on group presentations' were the teaching practices most supportive to developing convergent thinking.

Summary

Recent research has found that training programs or interventions can support improvements in creative outcomes, particularly students' ideation skills in divergent thinking tasks. These findings align with those of an earlier review of the effectiveness of creativity training, which found that well-designed creativity training programs typically do induce gains in performance that generalise across different creativity criteria (e.g. fluency, flexibility, originality), settings, and target populations (Scott, Leritz and Mumford, 2004[25]). Factors contributing to the relative effectiveness of such training programs include focusing on developing cognitive skills and the heuristics involved in their practical application using realistic exercises.

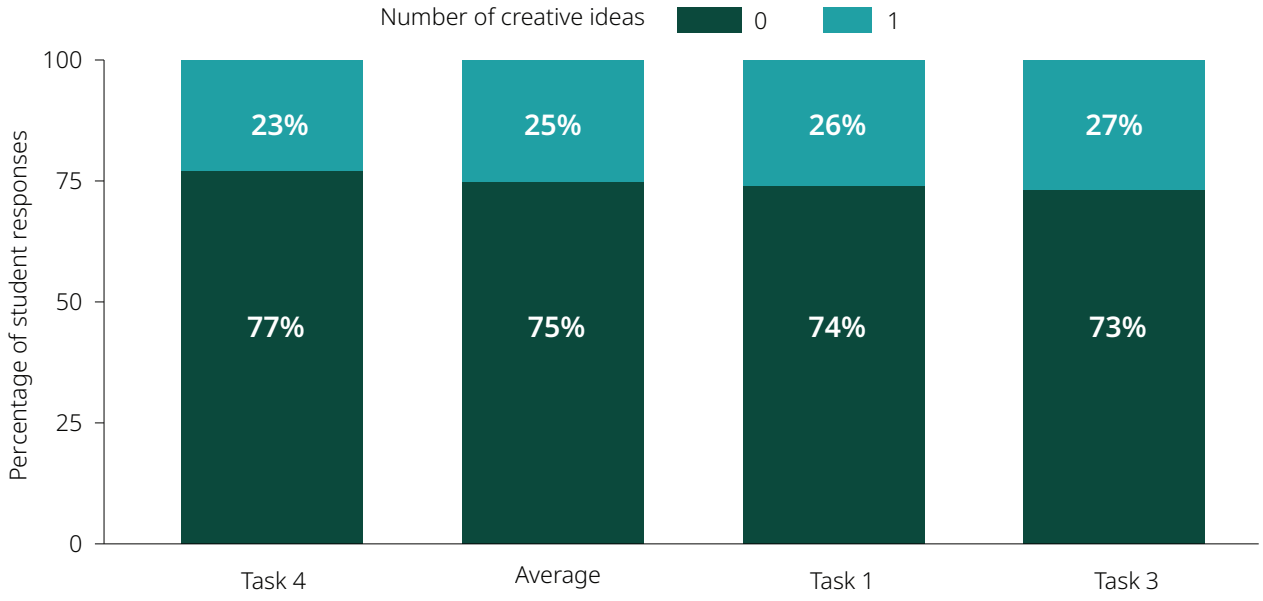
The school-based interventions described above also highlight how domain knowledge facilitates successful creative thinking outcomes, and the importance of developing well-designed, contextualised approaches to teaching creative thinking in educational settings. It may also be that such interventions typically require students to be more active participants in self-directed learning activities, thus increasing their motivation and metacognitive skills, which in turn facilitates their capacity to engage successfully in creative thinking in these contexts (Hu et al., 2013[26]).

2.2. Which ideation process led to more creative outcomes?

The findings in the first section of this chapter raise a related question: in which type of task on the PISA test were students more successful in thinking of creative ideas? In the 'generate creative ideas' tasks, students were explicitly asked to think of a single, original idea in response to the task stimulus. Perhaps surprisingly, students were far less likely to come up with a creative idea on these tasks compared to the divergent thinking tasks – with only around 25% of student responses on average across country-language groups scoring highly in appropriateness, originality and value in these tasks (Figure 2.3). This means that, on average across the different tasks, there is a 20-percentage point gap in the relative share of students that provided at least one creative idea when instructed to focus on idea originality compared to generating diverse ideas of any quality (Figure 2.4).

Figure 2.3. Only one quarter of responses to 'generate creative ideas' tasks were creative

Percentage of student responses that included a creative idea, on average across country-language groups, by task



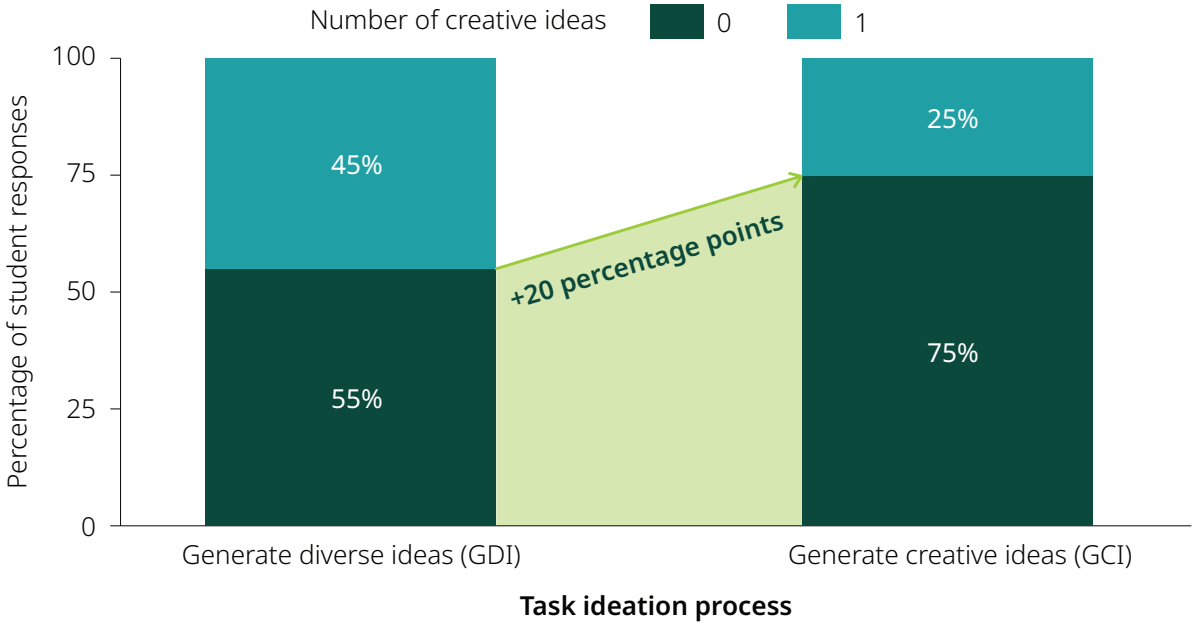
Note: Average percentages for each task were computed by country-language group then averaged across country-language groups. 'Creative ideas' are defined as those that scored at least 1 score point in appropriateness (out of a possible 2 score points) and that scored at least 2 score points in both originality and value (out of a possible 3 score points each).

Source: Table B2.3

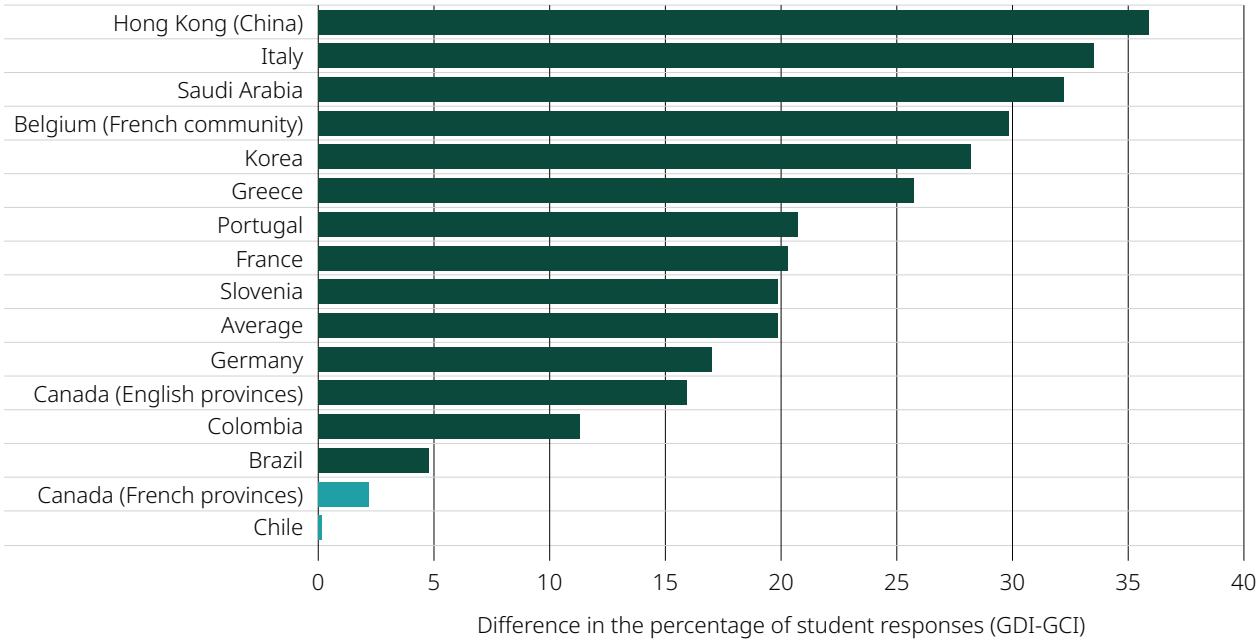
This trend is consistent across most country-language groups included in the PISA CT Rescoring study, although the relative shares of student responses with at least one creative idea on average across the two task types are similar in Brazil, Canada (English provinces) and Chile (Figure 2.4). Differences in the relative shares of responses that included at least one creative idea across the two task types are most pronounced in Hong Kong (China), Italy and Saudi Arabia, with a difference of over 30-percentage points in favour of creative outcomes in the 'generate diverse ideas' tasks.

Figure 2.4. Divergent thinking tasks consistently led to more student responses with creative ideas

Percentage of responses with at least one creative idea, on average across country-language groups and tasks by ideation process



Difference in the percentage of responses with at least one creative idea (GDI-GCI), on average across tasks by ideation process, by country-language group



Note: Average percentages for each task ideation process were computed first by country-language group on the pooled responses across relevant tasks, then averaged across country-language groups. All four GDI tasks are included in the GDI task average. 'Creative ideas' are defined as those that scored at least 1 score point in appropriateness (out of a possible 2 score points) and that scored at least 2 score points in both originality and value (out of a possible 3 score points each). In all country-language groups except Chile and Canada (French provinces), which are shown in a lighter tone, differences are statistically significant.

Source: Table B2.4

While these differences are substantial, one should take into consideration that the task difficulty and constraints across task types were not the same and therefore may have influenced the findings in our study. Nonetheless, the mean criteria scores of ideas across the two task types can shed additional light on the potential sources of these differences in creative outcomes (Table B2.2). On average across country-language groups, ideas in both types of task are similarly appropriate. The average idea in 'generate diverse ideas' tasks tended to be slightly less original than those in 'generate creative ideas' tasks, though the difference is small (less than 0.1 score points). This aligns with previous research that has found idea flexibility (the focus of the PISA divergent thinking tasks) to be more strongly associated with originality than other aspects of divergent thinking, for example fluency (Nijstad et al., 2010[11]). It may be that instructing students to focus on idea flexibility inevitably brings forth original ideas as a byproduct, regardless of whether originality is an explicit task demand. The criteria that is most different across the two task types is value, with ideas scoring 0.15 score points higher in the 'generate diverse ideas' tasks than in the 'generate creative ideas' tasks. Perhaps directing students' attention to originality leads them to largely ignore the value of valuable ideas in their idea generation process, or to search for original ideas at the expense of value. In other words, focusing students on originality too early during idea generation might limit the utility of their idea generation process and result in a less creative pool of ideas from which to further iterate.

These differences in creative success across the two task types may also point towards difficulties in convergent thinking, that is identifying, evaluating and selecting ideas to move forward. Research has found that individuals tend to have difficulties in recognising and selecting creative ideas, opting instead to choose more feasible and desirable ideas (Rietzschel, Nijstad and Stroebe, 2014[27]). In both types of PISA tasks, it's reasonable to assume that students engage in divergent thinking (i.e. thinking of multiple possible ideas), with students additionally needing to apply convergent thinking (i.e. choosing their most original idea) in the 'generate creative ideas' task. It may be at this point, rather than the idea generation phase, that students encounter difficulties. The data show that while the mean originality score of students' ideas in the 'generate creative ideas' tasks was 1.2 score points, the mean originality score of their most creative idea in the 'generate diverse ideas' tasks was much higher at 1.7 score points (a difference of around 0.5 score points) (Table B2.2 and Table B2.5). Although the different task instructions may have differentially constrained the quality of ideas that students generated and thus were able to choose from, it may also be that students did not successfully choose their most creative idea in the 'generate creative ideas' tasks. Further examining the role of convergent thinking in the creative process, the factors influencing its success (e.g. domain-knowledge, affective states), and instructional strategies to promote it is an area worthy of focus in future research (van Broekhaven, 2025[28]).

2.3. Are last ideas always better in an idea generation process?

As described above, student responses to the PISA tasks tended to be slightly more creative on average in the 'generate diverse ideas' tasks (mean idea sum score of 3.7 points) compared to the 'generate creative ideas' tasks (mean idea sum score of 3.6 points) (Table B2.2). These score differences become much starker when averaging students' 'best' ideas only in each task, with the mean 'best idea' sum scores in the 'generate diverse ideas' tasks totalling 5.4 score points on average across tasks (Table B2.2 and Table B2.5). Where do these 'best ideas' occur? Is it that students' last ideas in an idea generation process tend to be of higher quality?

Some empirical studies have found that later responses tend to be more creative than earlier ones in divergent thinking tasks (Benedek and Neubauer, 2013[29]; Benedek et al., 2014[30]; Christensen, Guilford and Wilson, 1957[31]; Piers and Kirchner, 1971[32]), whereas others have shown mixed results depending on the study conditions (Camarda et al., 2018[33]). Other researchers have also suggested that this order effect is mediated by various factors including fluid intelligence (Beaty et al., 2014[34]; Beaty and Silvia, 2012[35]) and executive control mechanisms (Beaty et al., 2014[34]; Ezzat et al., 2017[36]; Zabelina and Robinson, 2010[37]; Beaty et al., 2016[8]; Hao et al., 2015[38]) that inhibit the relevant but unoriginal ideas that typically come to mind at the beginning of a divergent thinking task (Gilhooly et al., 2007[39]).

We examined this idea order effect on the three 'generate diverse ideas' tasks in the PISA CT Rescoring study that require students to think of three different ideas. Our results show that, unlike prior research, students' first ideas tended to be their most creative idea and their last ideas least creative (Table 2.1). This general observation is also confirmed by looking at the frequency of students' most creative idea per response idea position, with well over half (58%) submitted as students' first idea on average across these three tasks (Table B2.7). This trend was also consistent across all country-language groups. Each of the individual criteria scores also tended to decrease as idea position increased, although idea originality remained relatively stable on average.

Table 2.1. First ideas in 'generate diverse ideas' tasks tended to be more creative

Mean idea sum and criteria scores by idea position in GDI tasks (*), on average across country-language groups

Response idea position	Appropriateness score	Originality score	Value score	Idea sum score
1	1.5	1.1	1.5	4.2
2	1.4	1.2	1.3	3.8
3	1.3	1.1	1.2	3.3

Note: (*) As Task 6 – *Save the River* requires only two ideas, rather than three ideas, the GDI task average excludes data from Task 6. Mean idea criteria scores for GDI tasks were computed first by country-language group on the pooled responses across relevant tasks, then averaged across country-language groups. Scores shown in the table are rounded to the nearest decimal point.

Source: Table B2.6.

One reason for this may be that the PISA tasks focus on idea flexibility, specifically, rather than fluency or remoteness, which is often the focus of other divergent thinking tasks. Given the limited number of ideas required in the PISA tasks, it might be that students already engage in some executive control mechanisms to select only their best ideas as part of their response, opting to submit their best idea first. Recent research has found that the number of ideas provided in a task can influence the relationship between originality and serial order effect, with the ideas of participants with lower fluency being both more original and elaborate than those with higher fluency regardless of idea position (Gonthier and Besançon, 2024[40]). There is also a dependency between students' ideas in the PISA tasks, given the instruction to provide ideas that are as different as possible to each other. Students may be able to think of creative ideas that are relatively similar, in which case they may choose to replace more creative ideas with more familiar yet different ones as the task progresses.

Key findings and implications

- Around 1 in 2 student responses to 'generate diverse ideas' tasks included at least one creative idea, on average. However, large differences were observed across country-language groups, with around 9 of every 10 student responses in Korea including at least one creative idea in these tasks compared to just 1 in 10 responses in Chile.
  - Idea generation activities with a focus on idea flexibility, even if limited in scope, can support creative idea generation in the context of everyday tasks.
- In general, a larger proportion of student ideas in the PISA 'generate diverse ideas' tasks (focused on idea flexibility) were considered creative than in the 'generate creative ideas' tasks (focused on idea originality). Only 1 in 4 responses to 'generate creative ideas' tasks included a creative idea (compared to around 1 in 2 in 'generate diverse ideas' tasks).
  - Constraining idea generation activities to original ideas only may serve to limit the quality of ideas that students are able to generate. Focusing on coming up with a few different ideas during the ideation phase may be a better strategy.
  - Students may also have difficulties in evaluating and selecting creative ideas (i.e. convergent thinking). Domain-relevant knowledge and skills support such decision-making processes by facilitating students' understanding of what constitutes a "good" creative idea.
- Experimental studies focused on developing creativity have largely focused on training divergent thinking skills, with findings generally supporting that these skills can be trained and that they are associated with more creative outcomes (see Box 2.2).
  - Successful training programs tend to focus on developing cognitive skills and include activities related to their practical application in real-world contexts.
  - Effective school-based interventions have focused on contextualised approaches to creative thinking, weaving creative thinking techniques with the acquisition of domain-specific knowledge.

References

Agogué, M. et al. (2014), "The impact of type of examples on originality: explaining fixation and stimulation effects", *Journal of Creative Behavior,* Vol. 48/1, pp. 1-12, https://doi.org/10.1002/jocb.37. [15]

Barbot, B. (ed.) (2016), "Creativity doesn't develop in a vacuum", *New Directions for Child and Adolescent Development,* Vol. 151, pp. 9-20, https://doi.org/10.1002/cad.20151. [9]

Barbot, B., M. Besançon and T. Lubart (2011), "Assessing Creativity in the Classroom", *Open Education Journal*, Vol. 4/1, pp. 58 - 66, https://doi.org/10.2174/1874920801104010058. [1]

Beaty, R. et al. (2016), "Creative Cognition and Brain Network Dynamics", *Trends in Cognitive Sciences*, Vol. 20/2, pp. 87-95, https://doi.org/10.1016/J.TICS.2015.10.004. [8]

Beaty, R. and P. Silvia (2012), "Why do ideas get more creative across time? An executive interpretation of the serial order effect in divergent thinking tasks.", *Psychology of Aesthetics, Creativity, and the Arts,* Vol. 6/4, pp. 309-319. [35]

Beaty, R. et al. (2014), "The roles of associative and executive processes in creative cognition", *Memory and Cognition*, Vol. 42/7, pp. 1186-1197, https://doi.org/10.3758/S13421-014-0428-8. [34]

Benedek, M. et al. (2014), "To create or to recall? Neural mechanisms underlying the generation of creative new ideas", *NeuroImage*, Vol. 88, pp. 125-133, https://doi.org/10.1016/J.NEUROIMAGE.2013.11.021. [30]

Benedek, M. and A. Neubauer (2013), "Revisiting Mednick's model on creativity-related differences in associative hierarchies. Evidence for a common path to uncommon thought", *Journal of Creative Behavior,* Vol. 47/4, pp. 273-289, https://doi.org/10.1002/JOCB.35. [29]

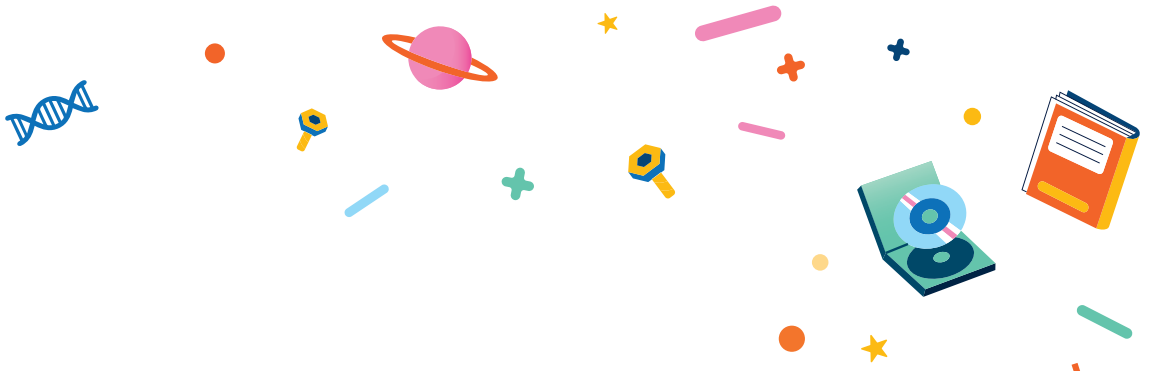
Camarda, A. et al. (2018), "Neural basis of functional fixedness during creative idea generation: An EEG study", *Neuropsychologia*, Vol. 118, pp. 4-12, https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2018.03.009. [33]

Christensen, P., J. Guilford and R. Wilson (1957), "Relations of creative responses to working time and instructions", *Journal of Experimental Psychology*, Vol. 53/2, pp. 82-88, https://doi.org/10.1037/H0045461. [31]

Corazza, G. and S. Agnoli (eds.) (2015), *Creative processes: Art, design and science,* Springer. [10]

Cropley, A. (2006), "In Praise of Convergent Thinking", *Creativity Research Journal,* Vol. 18/3, pp. 391-404, https://doi.org/10.1207/s15326934crj1803_13. [2]

De Jonge, K., E. Rietzschel and N. Van Yperen (2018), "Stimulated by novelty? The role of psychological needs and perceived creativity", *Personality and Social Psychology Bulletin,* Vol. 44/6, https://doi.org/10.1177/0146167217752236. [17]



Dygert, S. and A. Jarosz (2020), "Individual differences in creative cognition", *Journal of Experimental Psychology: General*, Vol. 149/7, pp. 1249-1274, https://doi.org/10.1037/xge0000713. [3]

Ezzat, H. et al. (2017), "How minimal executive feedback influences creative idea generation", *PLOS ONE*, Vol. 12/6, https://doi.org/10.1371/JOURNAL.PONE.0180458. [36]

Gilhooly, K. et al. (2007), "Divergent thinking: Strategies and executive involvement in generating novel uses for familiar objects", *British Journal of Psychology*, Vol. 98/4, pp. 611-625, https://doi.org/10.1111/J.2044-8295.2007.TB00467.X. [39]

Gonthier, C. and M. Besançon (2024), "It is not always better to have more ideas: Serial order and the trade-off between fluency and elaboration in divergent thinking tasks", *Psychology of Aesthetics, Creativity, and the Arts*, Vol. 18/4, pp. 480-492, https://doi.org/10.1037/aca0000485. [40]

Guilford, J. (1967), "Creativity: Yesterday, Today and Tomorrow", *The Journal of Creative Behavior,* Vol. 1/1, pp. 3-14, https://doi.org/10.1002/j.2162-6057.1967.tb00002.x. [4]

Guilford, J. (1956), "The structure of intellect", *Psychological Bulletin*, Vol. 53/4, pp. 267-293, https://doi.org/10.1037/H0040755. [5]

Hao, N. et al. (2015), "More mind wandering, fewer original ideas: Be not distracted during creative idea generation", *Acta Psychologica,* Vol. 161, pp. 110-116, https://doi.org/10.1016/J.ACTPSY.2015.09.001. [38]

Howard-Jones, P. (2002), "A dual-state model of creative cognition for supporting strategies that foster creativity in the classroom", *International Journal of Technology and Design Education,* Vol. 12/3, pp. 215-226, https://doi.org/10.1023/A:1020243429353. [12]

Hu, W. et al. (2013), "Increasing students' scientific creativity: The "Learn to Think" intervention program", *The Journal of Creative Behavior*, Vol. 47/1, pp. 3-21, https://doi.org/10.1002/jocb.20. [26]

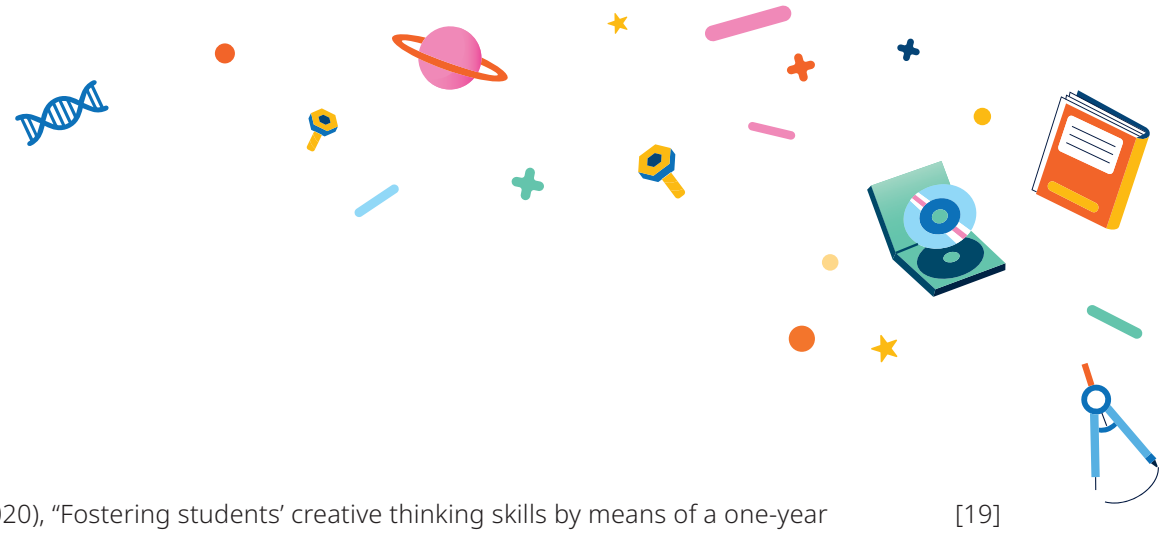
Meinel, M. et al. (2018), "Explorin the effects of creativity training on creative performance and creative self-efficacy: Evidence from a longitudinal study", *Journal of Creative Behavior*, Vol. 53/4, pp. 546-558, https://doi.org/10.1002/jocb.234. [22]

Nijstad, B. et al. (2010), "The dual pathway to creativity model: Creative ideation as a function of flexibility and persistence", *European Review of Social Psychology,* Vol. 21/1, pp. 34-77, https://doi.org/10.1080/10463281003765323. [11]

OECD (2023), *PISA 2022 Assessment and Analytical Framework*, OECD Publishing, Paris, https://doi.org/10.1787/dfe0bf9c-en. [13]

Piers, E. and E. Kirchner (1971), "Productivity and uniqueness in continued word association as a function of subject creativity and stimulus properties", *Journal of Personality,* Vol. 39/2, pp. 264-276, https://doi.org/10.1111/J.1467-6494.1971.TB00041.X. [32]

Rietzschel, E., B. Nijstad and W. Stroebe (2014), "Effects of problem scope and creativity instructions on idea generation and selection", *Creativity Research Journal*, Vol. 26/2, pp. 185-191, https://doi.org/10.1080/10400419.2014.901084. [27]



Ritter, S. et al. (2020), "Fostering students' creative thinking skills by means of a one-year creativity training program", *PLoS ONE*, Vol. 15/3, https://doi.org/10.1371/journal.pone.0229773. [19]

Ritter, S. and N. Mostert (2017), "Enhancement of creative thinking skills using a cognitive-based creativity training", *Journal of Cognitive Enhancement,* Vol. 1, pp. 243-253, https://doi.org/10.1007/s41465-016-0002-3. [20]

Scott, G., L. Leritz and M. Mumford (2004), "The effectiveness of creativity training: A quantitative review", *Creativity Research Journal*, Vol. 16, pp. 361-388. [25]

Smith, S., T. Ward and R. Finke (eds.) (1995), *The creative cognition approach*, The MIT Press. [7]

Sowden, P., A. Pringle and L. Gabora (2018), "The shifting sands of creative thinking: Connections to dual-process theory", in Gilhooly, K., L. Ball and L. Macchi (eds.), *Insight and Creativity in Problem Solving,* Routledge, https://doi.org/10.4324/9781315144061-3. [6]

Sun, M., M. Wang and R. Wegerif (2020), "Effects of divergent thinking training on students' scientific creativity: The impact of individual creative potential and domain knowledge", *Thinking Skills and Creativity*, Vol. 37, https://doi.org/10.1016/j.tsc.2020.100682. [23]

Vally, Z. et al. (2019), "Examining the effects of creativity training on creative production, creative self-efficacy, and neuro-executive functioning", *Creativity and Thinking Skills,* Vol. 71, pp. 70-78, https://doi.org/10.1016/j.tsc.2018.11.003. [21]

van Broekhaven, K. (2025), "The evaluation and selection of creative ideas in educational settings: Current knowledge and future directions", *Creativity Research Journal,* Vol. 37/2, pp. 307-131, https://doi.org/10.1080/10400419.2023.2253403. [28]

van Broekhoven, K. et al. (2020), "Fostering university students' idea generation and idea evaluation skills with a cognitive-based creativity training", *Creativity: Theories - Research - Applications,* Vol. 7/2, pp. 284-308, https://doi.org/10.2478/ctra-2020-0015. [18]

Vernon, D., I. Hocking and T. Tyler (2016), "An evidence-based review of creative problem solving tools: A practitioner's resource", *Human Resource Development Review,* Vol. 15/2, pp. 230-259, https://doi.org/10.1177/1534484316641512. [14]

Yang, K. et al. (2016), "Investigation of effective strategies for developing creativescience thinking", *International Journal of Science Education*, Vol. 38/13, pp. 2133-2151, http://dx.doi.org/10.1080/09500693.2016.1230685. [24]

Yuan, H. et al. (2021), "Examples facilitate divergent thinking: The effects of timing and quality", *Consciousness and cognition*, Vol. 93, https://doi.org/10.1016/j.concog.2021.103169. [16]

Zabelina, D. and M. Robinson (2010), "Creativity as flexible cognitive control", *Psychology of Aesthetics, Creativity, and the Arts*, Vol. 4/3, pp. 136-143, https://doi.org/10.1037/A0017379. [37]

03

Is creative thinking domain-specific, a more general skill – or both?

Can students be creative no matter what they do? Or is their creativity bound to specific areas of expertise? Understanding whether creativity is domain-general or domain-specific has been a subject of ongoing research over the past century (Miravete and Tricot, 2024[1]). Early research considered creative abilities to be influenced by a set of enduring personality traits and thinking patterns that support creativity across many kinds of tasks. More recently, research has shifted towards a domain-specific understanding of creativity, acknowledging that creative work in different domains draws upon different internal resources including knowledge, skills and attitudes. Some researchers argue that the domain-general vs. domain-specific dichotomy in creativity research is unhelpful as it obscures the interplay of both domain-general ("context-free") and domain-specific ("context-dependent") factors that support the creative thinking process (Plucker and Beghetto, 2006[2]).

The PISA 2022 assessment reported students' performance on the creative thinking test as part of a unidimensional scale, following the application of an Item-Response-Theory (IRT) model to students' test-item scores (OECD, 2024[3]). The IRT model indicated weakly correlated residuals between test items, supporting the construction of a unidimensional scale and suggesting that student performance on the PISA test largely captured an underlying latent creative thinking trait. Nonetheless, this latent scale obscured some differences in performance across task types and domain contexts. Item-level analyses found that students across countries and economies showed different strengths and weakness in creative thinking (OECD, 2024[4]). The PISA 2022 results lend support to the idea that creative thinking is supported by both context-free and context-dependent factors, both of which may be amenable to development through education.

What can the PISA 2022 Rescoring project data contribute to this debate? As described in the Introduction of this report, the PISA 2022 scoring method evaluated the extent to which students could successfully engage in idea generation, but it did not evaluate the holistic creativity of students' ideas nor identify the most creative responses within each country and economy. The PISA CT Rescoring project allows us to do this. To what extent are students capable of coming up with very creative ideas across different tasks and domain contexts? Who are these high-scoring creative all-rounders, and what factors might support consistent creative performance across different types of tasks?

3.1. Did students perform consistently across creative thinking tasks?

Of the seven tasks included in the PISA 2022 Rescoring project, students in the sample completed between two and four tasks each (see Annex A1). Table 3.1 presents the correlations of residuals between the seven tasks included in the study, on average across country-language groups, illustrating how student scores across items relate after accounting for a shared underlying trait. The generally weak correlations support the idea that performance in the PISA test was driven primarily by a single dimension (i.e. skill in creative thinking) rather than task- or domain-specific factors. Nonetheless, some weak clustering between tasks is apparent – particularly within the problem-solving tasks asking students to generate multiple ideas (Tasks 5 to 7; see Figures B-H in the Introduction for a presentation of the tasks included in the study). This may be expected to some extent, as the task demands and format in this group (generate one or more solution ideas of any quality) are more similar than those of the remaining task types. Indeed, task design and instructions have been found to affect performance and cognitive processes in creative thinking tasks (Acar and Shen, 2025[5]).

Another way to consider the extent to which students can think creatively across different tasks is to examine their success at the individual level. The PISA CT Rescoring project sample included 5530 students who completed at least three different tasks (around 300-400 students per country-language group) (Annex A1). Figure 3.1 shows that while just over half of all students (52%) that sat at least three tasks were high-scoring in one or two of them – with "high-scoring" defined as scoring in the top quartile of the standardised holistic creativity scores relative to their country-language-task group peers (see Box 3.1) – only 8% of students were high-scoring in three or more tasks. Moreover, over 1 in 3 students were not able to come up with a single high-scoring response relative to their country-language group peers in any of the tasks included in the study.

This relatively low share of "creative all-rounders" across country-language groups – defined as students who achieved at least three high-scoring responses (see Box 3.1) – might be affected by the types of tasks included in our study. The four tasks classified as 'generate diverse ideas' tasks only instructed students to provide ideas that were as *different* to each other as possible, but to achieve a high holistic creativity score, students may have needed to suggest one or more *creative* ideas within the same response. The relative ranking of the holistic scoring method mitigates this issue to some extent, as responses are only compared to others in the same country-language-task group, but these task types may nonetheless place relatively high demands on students and thus make it harder to consistently achieve high-scoring responses across tasks.

Table 3.1. Average residual correlations between tasks included in the PISA CT Rescoring study

	Creative expression tasks			Creative problem-solving tasks			
	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7
Task 1 Science Fair Poster	1	0.04	-0.02	**-0.09**	- 0.04	-0.06	0.02
Task 2 Illustration Titles		1	**0.06**	**0.04**	**0.07**	**0.06**	**0.13**
Task 3 2983			1	0.03	**0.11**	**0.11**	-0.01
Task 4 Save the Bees				1	**0.07**	**0.06**	**0.15**
Task 5 Food Waste					1	**0.10**	**0.17**
Task 6 Save the River						1	**0.12**
Task 7 The Exhibit							1

Note: Bolded values show significant correlations at the average country-language group level. A Generalized Partial Credit Model (GPCM) was modelled for each country-language group, using the (unstandardised) holistic creativity scores. Higher values of residual correlation can be interpreted as evidence of other shared factors, beyond a shared underlying creative thinking skill, that influences performance in task pairs.

Source: Table B3.1

Box 3.1. Which score data are used in this chapter?

The analyses in this chapter mainly draw upon the **holistic creativity scores** (Annex A2). Each student response was scored based on judges' evaluations of the overall creative quality of the response. The holistic creativity score reflects a relative ranking of creative quality compared to all other responses in that country-language-task group. To account for differences in the distribution of scores across country-language-task groups, the analyses in this chapter **standardised and transformed** the raw holistic creativity scores onto a 1–10-point scale, with a mean of 5 score points.

Terms used in this chapter

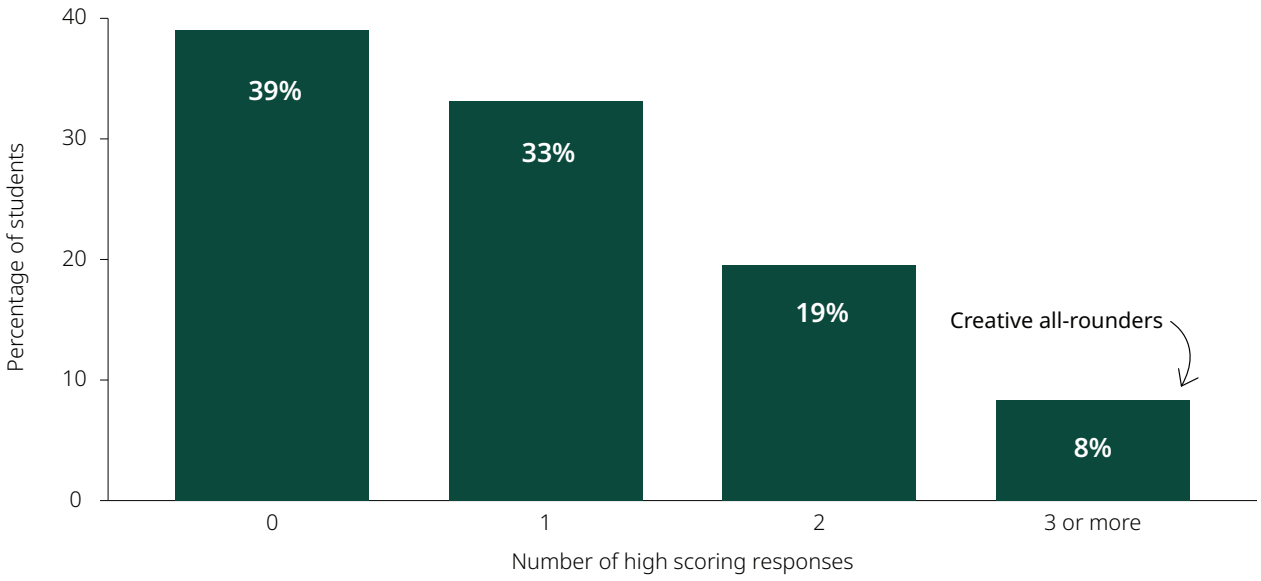
- **Jittered standardised holistic score**: For some of the analyses in this chapter, the standardised holistic creativity scores were jittered. This involves adding a small random error, to the magnitude of 1e-6, onto each standardised holistic creativity score. The jittered scores ensure that there is an even distribution of students across quantiles.

- **High-scoring answer**: A student response was classified as high scoring if its jittered standardised holistic score fell within the top quartile of scores (i.e. at or above the 75th percentile) relative to all other responses within its country-language-task group. Note that the creative quality of high-scoring answers across country-language groups may vary due to the relative ranking nature of the holistic scoring method.
- **Creative all-rounders**: Students that responded to three or four tasks in the PISA CT Rescoring project and who achieved at least three high-scoring answers. Students with three or four high-scoring answers were considered together given the small sample sizes (see Annex A1).

This chapter focuses on examining the relative shares of creative all-rounders across country-language groups and understanding the factors that influence students' capacity to consistently achieve high holistic creativity scores across different types of task.

Figure 3.1. Few students across country-language groups manage to be creative all-rounders

Share of students that score in the top quartile of the standardised holistic creativity scores in multiple tasks (on average across country-language groups)



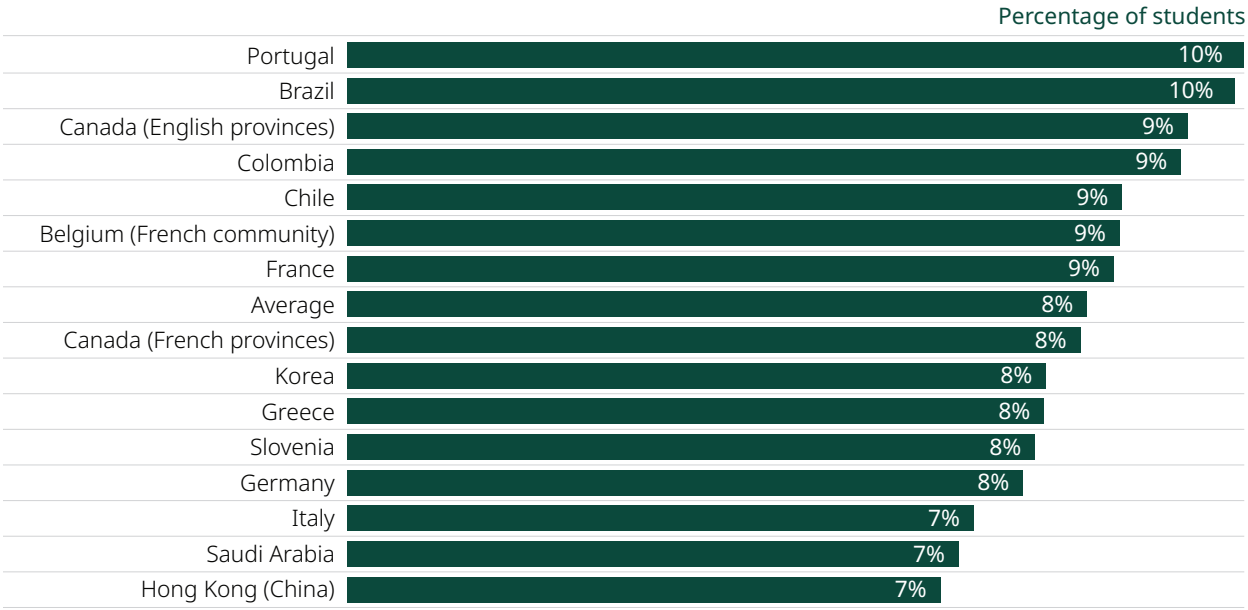
Note: Sample includes all students who took at least 3 tasks included in the PISA CT Rescoring project (N = 5530). Results shown are at the average country-language group level and are rounded to the nearest integer. "High scoring" responses are jittered standardised holistic scores that fall within the top quartile (75th percentile) of each country-language-task group. The standardised holistic creativity scores were jittered, i.e. a random small digit (1e-6) was added to each score, so that equal quartiles in each country-language group could be produced. Results shown are the average of 20 replications of the analysis to account for sampling and jitter-induced variability.

Source: Table B3.2

Figure 3.2 shows that the share of creative all-rounders within each country-language group varies slightly, with around 10% of students in Portugal and Brazil achieving high-scoring responses compared to their country-language peers across three or more tasks, compared to around 7% of students who performed well across multiple tasks in Italy, Saudi Arabia and Hong Kong (China). In countries with fewer creative all-rounders compared to the country-language average, there is a more varied spread of students that come up with creative ideas across the different tasks in the study.

Figure 3.2. Some country-language groups have more creative all-rounders than others

Percentage of students in each country-language group that achieved three or more high-scoring responses (relative to their country-language-task group)



Note: Sample includes all students who took at least 3 tasks included in the PISA CT Rescoring project (N = 5530). Value labels shown in the figure are rounded to the nearest integer. "High scoring" responses are jittered standardised holistic scores that fall within the top quartile (75th percentile) of each country-language-task group. The standardised holistic creativity scores were jittered, i.e. a random small digit (1e-6) was added to each score, so that equal quartiles in each country-language group could be produced. Results shown are the average of 20 replications of the analysis to account for sampling and jitter-induced variability.

Source: Table B3.2

3.2. Who are the creative all-rounders and what characteristics do they share?

Students who achieved high holistic creativity scores across three or more tasks were relatively rare, representing less than 1 in 10 students on average across country-language groups and task combinations. What factors could account for the variation in student performance across tasks and differences in the proportion of creative all-rounders across country-language groups?

3.2.1. Students' socio-demographic background (gender, socio-economic status) and academic performance

Table 3.2 shows the share of boys and girls amongst students who provided different numbers of high-scoring responses. The share of female students increases within each group as the number of high-scoring responses increases, with the gender gap being largest amongst the creative all-rounders group (59% of creative all-rounders are girls compared to 41% boys, on average across country-language-task groups). As reported in the PISA 2022 results (OECD, 2024[4]) and explored later in Chapter 6 of this report, girls tended to perform better in the PISA 2022 creative thinking tasks overall – so finding that they also managed to achieve high-scoring responses more consistently than boys is no surprise.

Table 3.2. More girls than boys are creative all-rounders

Percentage of girls and boys (on average across country-language groups), by number of high-scoring responses

High-scoring answers	Share of girls	Share of boys	Difference (Girls-boys)
0	47%	53%	**-6%**
1	49%	51%	-2%
2	54%	46%	**9%**
3 or more	59%	41%	**18%**

Note: Significant values are highlighted in bold. Percentages are rounded to the nearest integer. Sample includes all students who took at least 3 tasks included in the PISA CT Rescoring project (N = 5530). "High scoring" responses are jittered standardised holistic scores that fall within the top quartile (75th percentile) of each country-language-task group. The standardised holistic creativity scores were jittered, i.e. a random small digit (1e-6) was added to each score, so that equal quartiles in each country-language group could be produced. Results shown are the average of 20 replications of the analysis to account for sampling and jitter-induced variability. Results shown are at the average country-language group level.

Source: Table B3.3

Table 3.3 similarly shows that socio-economic advantage – which is typically associated with higher scores in PISA, in general, but also found to be positively associated to performance in creative thinking (OECD, 2024[4]) – is associated with students' capacity to consistently achieve high-scoring responses across different types of tasks. On average across country-language groups, around 4 in 10 creative all-rounders were amongst the top quartile of their country-language socio-economic index, while only 1 in 10 creative all-rounders were in the lowest quartile of their country-language socio-economic index.

In addition to students' background (gender and socio-economic status), students' performance on reading and mathematics likely influence the probability that they are a creative all-rounder. Indeed, the PISA 2022 results showed that differences in the performance of advantaged and disadvantaged students on the PISA test were largely accounted for (albeit not completely) by students' mathematics and reading scores (OECD, 2024[4]). Unsurprisingly, the mean mathematics and reading scores of students increases as the number of high-scoring responses increases, on average across task combinations and country-language groups (Table B3.5). The mean score difference between students with no high-scoring answers and students that are creative all-rounders is more than the mean standard deviation (OECD countries) for both mathematics and reading in PISA 2022. These differences were statistically significant for both subjects in all country-language groups.

Table 3.3. More advantaged students than disadvantaged students are creative all-rounders

Percentage of students in the top / bottom quartile of the country-language socio-economic index (on average across country-language groups), by number of high-scoring responses

High-scoring answers	Advantaged students	Disadvantaged students	Difference (Disadvantaged-advantaged)
0	16%	29%	**12%**
1	24%	21%	-3%
2	33%	15%	**-18%**
3 or more	41%	10%	**-31%**

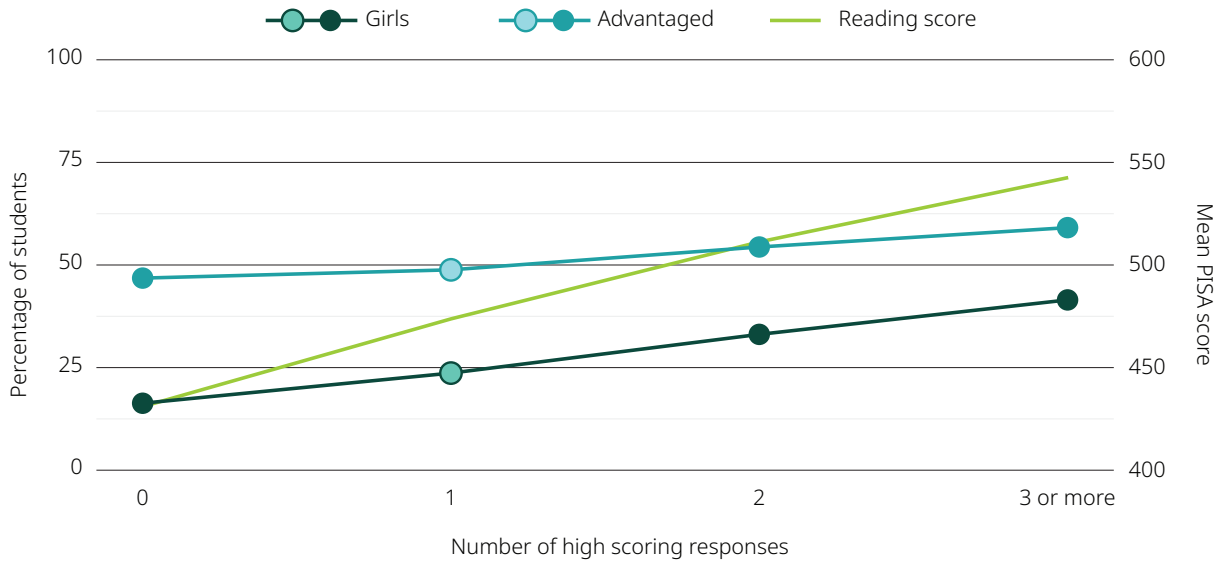
Note: Significant values are highlighted in bold. Percentages are rounded to the nearest integer. Sample includes all students who took at least 3 tasks included in the PISA CT Rescoring project (N = 5530). "High scoring" responses are jittered standardised holistic scores that fall within the top quartile (75th percentile) of each country-language-task group. The standardised holistic creativity scores were jittered, i.e. a random small digit (1e-6) was added to each score, so that equal quartiles in each country-language group could be produced. Results shown are the average of 20 replications of the analysis to account for sampling and jitter-induced variability. Results shown are at the average country-language group level.

Source: Table B3.4

Figure 3.3 summarises the associations between gender, socio-economic status and academic performance, and the number of high-scoring responses achieved by students. While these factors are all significantly associated with differences between being a creative all-rounder and achieving no high-scoring responses, how important are they in determining whether or not students score highly across tasks in the PISA test? Only around 20% of the variance in high-scoring responses is accounted for by these factors, on average across task combinations and country-language groups, suggesting that there are other important factors influencing students' capacity to consistently score highly across creative tasks (Table B3.6).

Figure 3.3. Creative all-rounders are more likely to be girls, be from an advantaged socio-economic background, and score more highly in reading than non-all-rounders

Share of girls / advantaged students (left y-axis) and mean PISA reading score (right y-axis), on average (across country-language groups), by number of high-scoring responses



Note: Significant differences between girls and boys, and advantaged and disadvantaged students, within each high-scoring response group are shown by markers in a darker shade. Sample includes all students who took at least 3 tasks included in the PISA CT Rescoring project (N = 5530). Advantaged students refer to students in the top quartile of their country-language ESCS index. "High scoring" responses are jittered standardised holistic scores that fall within the top quartile (75th percentile) of each country-language-task group. The standardised holistic creativity scores were jittered, i.e. a random small digit (1e-6) was added to each score, so that equal quartiles in each country-language group could be produced. Results shown are the average of 20 replications of the analysis to account for sampling and jitter-induced variability. Results shown are at the average country-language group level.

Source: Tables B3.3-B3.5

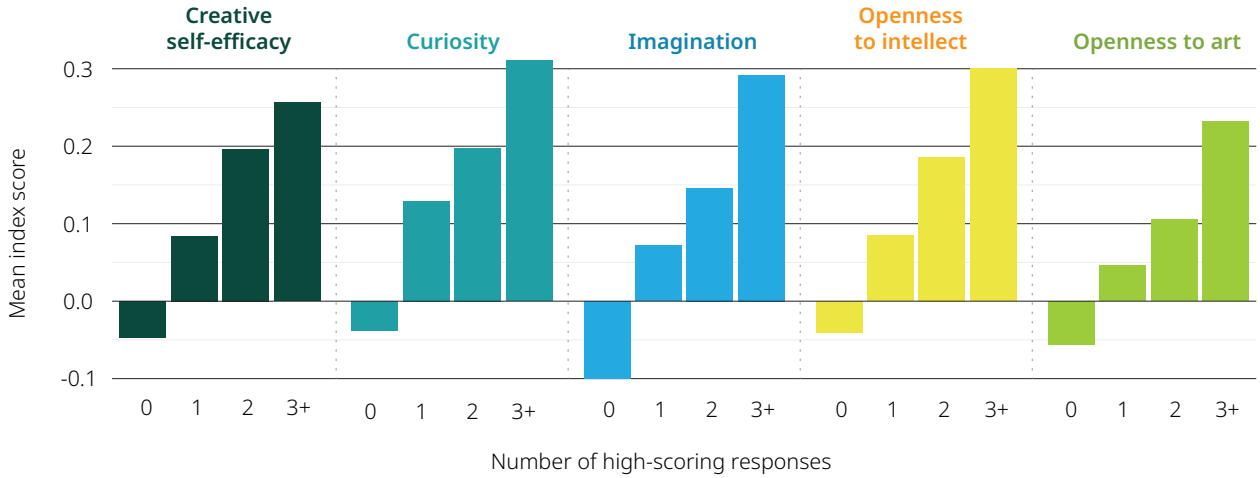
3.2.2. By personality traits and attitudes

Certain enduring personality traits, like openness and extraversion, have typically been associated with creative individuals (Puryear, Kettler and Rinn, 2017[6]). Other attitudes, like seeing oneself as imaginative or creative self-efficacy, have also been found to support creative outcomes in prior research (Abdullah, Omar and Panatik, 2016[7]). Some authors argue that these traits and attitudes can support creative work across domains (Van Broekhoven, Cropley and Seegers, 2020[8]; Miravete and Tricot, 2024[1]).

Figure 3.4 shows that creative all-rounders do generally hold more positive beliefs about their capacity to engage in creative tasks (creative self-efficacy), and they also see themselves as more curious, more imaginative, and more open to intellect and the arts than their peers – particularly those who did not come up with any or only one high-scoring response across tasks. Certain socio-emotional skills might also support students' performance in creative thinking tasks, for example, by supporting students to continue working on open-ended problems that require multiple responses. Figure 3.5 shows that creative all-rounders also tended to self-report higher levels of assertiveness and perseverance, but interestingly, much lower stress resistance than all other groups of students.

Figure 3.4. Creative all-rounders report stronger attitudes and beliefs associated with creativity

Mean index of creative attitudes (on average across country-language groups), by number of high-scoring responses



Note: Sample includes all students who took at least 3 tasks included in the PISA CT Rescoring project (N = 5530). Some scales from the PISA 2022 student questionnaire were suppressed from some country-language groups, and therefore they are not included in the average country-language group results; for more information, see the PISA 2022 Technical Report (OECD, 2024[3]). "High scoring" responses are jittered standardised holistic scores that fall within the top quartile (75th percentile) of each country-language-task group. The standardised holistic creativity scores were jittered, i.e. a random small digit (1e-6) was added to each score, so that equal quartiles in each country-language group could be produced. Results shown are the average of 20 replications of the analysis to account for sampling and jitter-induced variability.

Source: Table B3.7

Coming up with creative answers across several tasks in the PISA test requires that students remain engaged with the test and sustain a certain level of investment throughout; it may be that those who are less resistant to stress are more inclined to put in the effort to perform well, in order to reduce feelings of anxiety. For example, while some affective states might be considered as "deactivating states" that can decrease arousal and negatively impact creativity, stress – although triggered by negative emotions – might be considered an "activating state" for creativity that serves to enhance creative engagement by its association with persistence (Nijstad et al., 2010[9]; De Dreu and Nijstad, 2008[10]). The optimal level of arousal model also suggests that overstimulation from one's environment might promote an increase in creative thinking skills aimed at creating simpler input stimuli (i.e. convergent thinking), whereas understimulation might lead to inverse effects and rather promote an increase in exploratory behaviours and cognitive processes aimed at novelty and increasing complexity in one's environment (i.e. divergent thinking) (Gustafsson, 2023[11]). Several studies have investigated the relationship between mood, affect and creativity, with the varied findings across studies demonstrating the complex interactions between an individual's environment, personality and cognitive preferences (Baas, De Dreu and Nijstad, 2008[12]).

Figure 3.5. Creative all-rounders report higher levels of assertiveness and perseverance, but lower levels of stress resistance

Mean index of socio-emotional skills (on average across country-language groups), by number of high-scoring responses

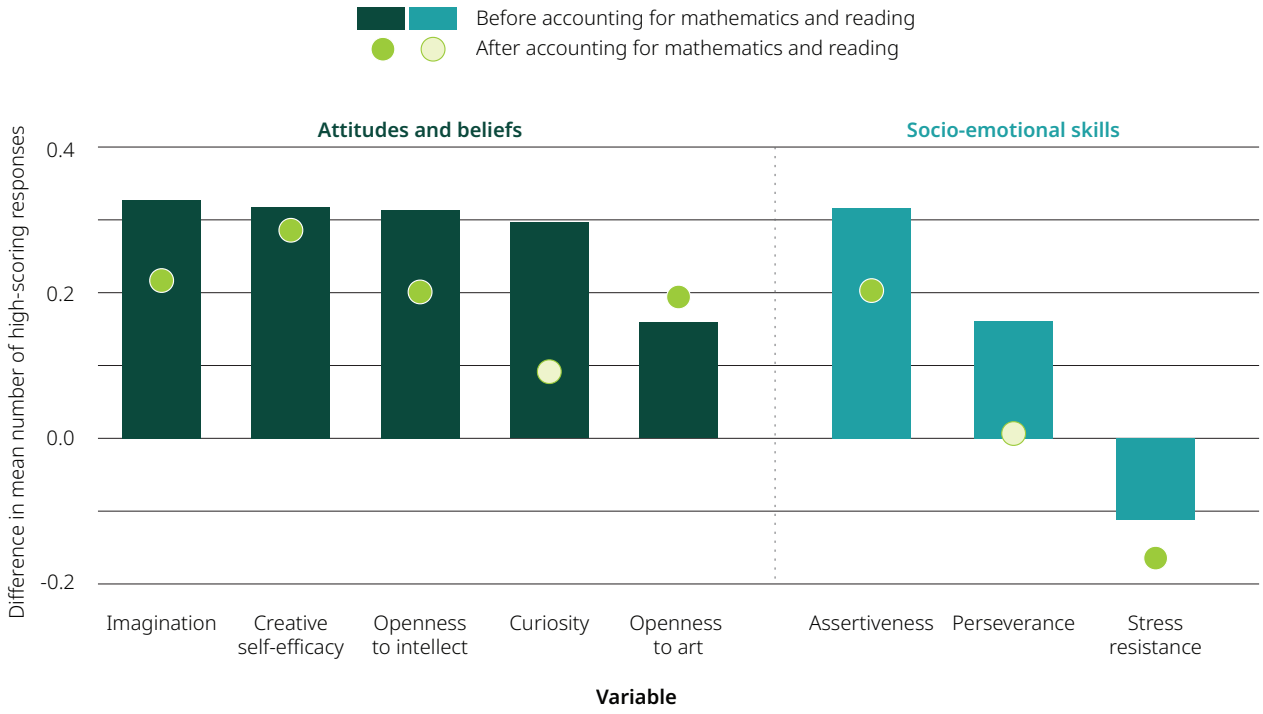
Note: Sample includes all students who took at least 3 tasks included in the PISA CT Rescoring project (N = 5530). Some scales from the PISA 2022 student questionnaire were suppressed from some country-language groups, and therefore they are not included in the average country-language group results; for more information, see the PISA 2022 Technical Report (OECD, 2024[3]). "High scoring" responses are jittered standardised holistic scores that fall within the top quartile (75th percentile) of each country-language-task group. The standardised holistic creativity scores were jittered, i.e. a random small digit (1e-6) was added to each score, so that equal quartiles in each country-language group could be produced. Results shown are the average of 20 replications of the analysis to account for sampling and jitter-induced variability.

Source: Table B3.8

Although several attitudes and socio-emotional skills clearly demonstrate a positive association with the number of high-scoring responses that students achieve, are these factors specific to supporting students' creative thinking performance or do they rather reflect a more general association with student performance in PISA? After accounting for students' mathematics and reading scores on the PISA test, several of the beliefs, attitudes and socio-emotional skills discussed in this chapter remain small yet significant predictors of being a creative all-rounder. However, they each explain only a marginal proportion of the variation in students' number of high-scoring responses across tasks before accounting for students' mathematics and reading scores, compared to the 17-18% of the variation explained after accounting for mathematics and reading (Table B3.9 and Table 3.10). Comparing the difference in the mean number of high-scoring responses between students in the top and bottom quarter of the self-reported indices, even after accounting for mathematics and reading scores, highlights the associations between all-round performance in creative thinking and certain attitudes and beliefs (Figure 3.6).

Figure 3.6. Students that report strong attitudes and socio-emotional skills associated with creativity are more likely to score highly across creative tasks

Difference in the mean number of high-scoring responses between students in the top / bottom quartile of select attitudinal and socio-emotional indices (on average across country-language groups), before and after accounting for mathematics and reading scores



Note: Bars show differences (top quartile - bottom quartile) before accounting for mathematics and reading, and markers show differences after accounting for mathematics and reading. Significant differences after accounting for mathematics and reading are shown by filled markers. All differences before accounting for mathematics and reading are significant. Sample includes all students who took at least 3 tasks included in the PISA CT Rescoring project (N = 5530). Some scales from the PISA 2022 student questionnaire were suppressed from some country-language groups, and therefore they are not included in the average country-language group results; for more information, see the PISA 2022 Technical Report (OECD, 2024[3]). "High scoring" responses are jittered standardised holistic scores that fall within the top quartile (75th percentile) of each country-language-task group. The standardised holistic creativity scores were jittered, i.e. a random small digit (1e-6) was added to each score, so that equal quartiles in each country-language group could be produced. Results shown are the average of 20 replications of the analysis to account for sampling and jitter-induced variability.

Source: Table B3.9-B3.10

Key findings and implications

- Many students that took the PISA test demonstrated that they are capable of being creative, with around 6 in 10 students scoring highly in at least one task, relative to their country-language group peers.
  - Successful creative thinking, on a task-by-task basis, does not appear limited to a particular student group.
- Few students in all country-language groups, and only around 8% of students on average, are creative all-rounders, i.e. students that consistently score highly in creative thinking tasks. Consistent creative performance across different types of tasks therefore appears much more exclusive and difficult to achieve than successful creative performance on individual tasks.
  - Students' gender and socio-economic status, and mathematics and reading performance were all significantly associated with scoring highly across multiple tasks, together accounting for close to 20% of the total variation in high-scoring outcomes.
  - Students who believed in their own capacity to be creative, who saw themselves as imaginative, open and curious, and who reported higher perseverance and assertiveness, also tended to achieve more high-scoring responses than those who did not.
- Our results are consistent with theories that suggest creative thinking is supported by both domain-general ("context-free") and domain-specific ("context-specific") factors (Plucker and Beghetto, 2006[2]).
  - While "context-free" factors include socio-demographic background, academic performance, and attitudes and beliefs, much of the variation in consistent creative thinking performance remains unexplained by these factors.
  - It may be that other general factors support performance across domains, like general intelligence or conscientiousness, or that context-specific factors like domain knowledge and skills or opportunities to practice creative thinking are indeed more important for supporting consistent performance and account for the rest of the variation in creative all-round performance.
  - Teachers and school environments can play a key role in fostering both types of supporting factors, for example by providing positive experiences that reinforce student beliefs about creativity, by developing their knowledge and skills, and by enabling them to practice applying creative thinking in domain-relevant ways.

References

Abdullah, I., R. Omar and S. Panatik (2016), "A literature review on personality, creativity and innovative behavior", *International Review of Management and Marketing*, Vol. 6/1, pp. 177-182. [7]

Acar, S. and Y. Shen (2025), "Along the convergent–divergent continuum: The role of task structure in the PISA Creative Thinking assessment", *The Journal of Creative Behavior*, Vol. 59/2, https://doi.org/10.1002/JOCB.70029. [5]

Baas, M., C. De Dreu and B. Nijstad (2008), "A meta-analysis of 25 years of research on mood and creativity: Hedonic tone, activation, or regulatory focus?", *Psychological Bulletin*, Vol. 134/6, pp. 739-756, https://doi.org/10.1037/a0012815. [12]

De Dreu, C. and B. Nijstad (2008), "Conflict and creativity: Threat-rigidity or motivated focus?", *Journal of Personality and Social Psychology*, Vol. 95, pp. 648-661, https://doi.org/10.1037/0022-3514.95.3.648. [10]

Gustafsson, E. (2023), "How can contextual variables influence creative thinking? Contributions from the optimal-level of arousal model", *Journal of Creative Behavior,* Vol. 57/1, pp. 96-108, https://doi.org/10.1002/jocb.565. [11]

Miravete, S. and A. Tricot (2024), "Are some people generally more creative than others? A systematic review of fifty years' research", *Educational Psychology Review 2024 36:4,* Vol. 36/4, pp. 1-39, https://doi.org/10.1007/S10648-024-09926-6. [1]

Nijstad, B. et al. (2010), "The dual pathway to creativity model: Creative ideation as a function of flexibility and persistence", *European Review of Social Psychology,* Vol. 21/1, pp. 34-77, https://doi.org/10.1080/10463281003765323. [9]

OECD (2024), *PISA 2022 Results (Volume III): Creative Minds, Creative Schools,* PISA, OECD Publishing, https://doi.org/10.1787/765ee8c2-en. [4]

OECD (2024), *PISA 2022 Technical Report,* PISA, OECD Publishing, https://doi.org/10.1787/01820d6d-en. [3]

Puryear, J., T. Kettler and A. Rinn (2017), "Relationships of personality to differential conceptions of creativity: A systematic review", *Psychology of Aesthetics, Creativity, and the Arts,* Vol. 11/1, pp. 59-68, https://doi.org/10.1037/ACA0000079. [6]

Sternberg, R., E. Grigorenko and J. Singer (eds.) (2006), *Why creativity is domain general, why it looks domain specific, and why the distinction does not matter.*, American Psychological Association, https://doi.org/10.1037/10692-009. [2]

Van Broekhoven, K., D. Cropley and P. Seegers (2020), "*Differences in creativity across Art and STEM students: We are more alike than unalike*", https://doi.org/10.1016/j.tsc.2020.100707. [8]

04

Are there cross-cultural differences in creative thinking?

Creativity is generally understood as resulting from the interaction among an individual's aptitudes, process and their *environment*, with the outcome being defined as both novel and useful *within a given social context* (Plucker, Beghetto and Dow, 2004[1]). Any creative product is therefore embedded within a particular social context, that in turn is inherently shaped by cultural norms and expectations as well as other socio-environmental factors.

Cultural norms and expectations can influence the skills that individuals choose to develop, the values that shape their personality development, and differences in performance expectations within societies (Niu and Sternberg, 2003[2]; Wong and Niu, 2013[3]; Lubart, 1998[4]; Shao et al., 2019[5]). For example, Rudowicz (2003[6]) analysed differences in creativity perceptions between "Western" and "Eastern" cultures noting that, in the West, creativity often implies a break with tradition and is primarily valued for solving problems or achieving personal success, whereas in the East, an individual's creativity is primarily valued for its social and moral contribution. Some studies have also investigated how cultural differences affect national measures of creativity and innovation, with differences along the individualism-collectivism spectrum significantly shaping how creative products can be defined and valued (Rinne, Steel and Fairweather, 2013[7]; Ng, 2003[8]).

Confluence approaches of creativity also emphasise the role that socio-environmental factors play in influencing an individual's capacity to engage in creative work (Amabile, 1983[9]; 2012[10]; Amabile and Pratt, 2016[11]; Sternberg and Lubart, 1991[12]; 1995[13]; Sternberg, 2006[14]). Findings from organisational research have demonstrated that practices and extrinsic motivators like receiving informal feedback, appropriate recognition and encouragement, engaging in goal setting, working as a team, and enjoying task autonomy all enable individuals to develop creative ideas (Amabile, 2012[10]; Zhou and Su, 2010[15]).

From an education perspective, cultural norms will inevitably influence the outcomes an education system values for its students and the content it prioritises in the curriculum. In some cases, educational approaches might actively discourage creative thinking and achievement at school (Wong and Niu, 2013[3]). For example, the

pressures of standardisation and accountability in educational testing systems might reduce opportunities for creative thinking in schoolwork (DeCoker, 2000[16]). Classroom practices will also be affected by their school environment. For example, school leaders and teachers can cultivate environments that help students learn when and how creative thinking is appropriate, encourage students to set their own goals and identify promising ideas, and take responsibility for contributing to creative teamwork (Beghetto and Kaufman, 2010[17]; 2014[18]). Certain classroom practices might also stifle creative thinking, for example by perpetuating the idea that there is only one way to learn or solve problems, by cultivating attitudes of submission and fear of authority, by promoting beliefs that creativity is a rare innate quality, or by discouraging students' curiosity and inquisitiveness in formal learning contexts (Nickerson, 2010[19]).

In Chapter 1 of this report, we explored what makes ideas creative and how the different score criteria used in this study (appropriateness, originality and value) relate to judges' holistic creativity scores. In this chapter, we focus on differences in how creativity is understood, manifested and evaluated across the different socio-linguistic groups included in the PISA CT Rescoring study. For example, what are the differences in terms of the quality of students' ideas across country-language groups? Is the extent to which the criteria of appropriateness, originality and value determine holistic creativity scores similar across socio-linguistic contexts?

The goal of this chapter is not to "re-rank" country-language groups by students' creative thinking performance. For a variety of reasons – including the limited subset of country-language groups, tasks and student data included in the study, as well as differences between the scoring methods involved – direct comparisons of student scores and the relationships between the different scoring methods across country-language groups should be interpreted with caution (see Box 4.1). Nonetheless, these data can still provide interesting insights that, when considered together as a whole, demonstrate that students and/or judges may differ in how they interpret "creative quality" across socio-linguistic contexts.

Box 4.1. Caution on interpreting the data in this chapter

Results reported in this chapter should be interpreted bearing in mind three key points: (1) the potential effects of judge bias in human scoring approaches; (2) the principles of the two scoring methods involved (holistic and criteria-based methods); and (3) the analytical approach used to compare criteria-based and holistic creativity scores.

Human scoring approaches: addressing rater bias and inter-rater reliability

Any human scoring approach is susceptible to rater bias. Several studies have demonstrated the effects of rater expertise (Kaufman et al., 2008[20]; Kaufman, Baer and Cole, 2009[21]; Long, 2014[22]) and subjective and socio-cultural bias (Lloyd-Cox, Pickering and Bhattcharya, 2022[23]; Kaufman et al., 2010[24]) on scoring creative tasks. The PISA CT Rescoring project implemented several measures intended to reduce rater bias, including anchor training exercises, inter-rater reliability checks, and using task-specific rubrics for attributing the criteria-based scores (see Annex A2). Nonetheless, differences in student scores across country-language groups may reflect differences in judge leniency and the relative importance they assign to different elements of the response (e.g. originality vs relevance), as much as they reflect differences in students' relative strengths or weaknesses or capacity to generate creative ideas.

Principles of the holistic and criteria-based scoring methods: relative ranking versus absolute scores

The two scoring methods referred to in this chapter differ in how judges assign scores (Box 4.2). The holistic creativity scores reflect a relative ranking of student responses in each country-language-task group, whereas the criteria-based scores reflect the quality of student ideas according to a centralised, task-specific rubric (see Annex A2). Given these key differences, ideas in a holistic score category in one country-language-task group may not be equivalent in terms of quality to ideas in the same holistic score category in another country-language-task group. These fundamental differences between the two scoring approaches should therefore be kept in mind when comparing the relative relationship between the two types of scores, both within and across country-language groups.

Analytical approach: averaging criteria-based scores across ideas in 'generate diverse ideas' tasks

The analytical approach adopted in this chapter makes an important assumption that each idea within a student response contributes equally to a judge's holistic creativity score for that response. For tasks that require students to propose multiple ideas (i.e. tasks classified as 'generate diverse ideas' items), idea criteria scores are first averaged within each response before being associated to the response holistic creativity score. However, it may be that some ideas within a response are more important than others for determining judges' holistic creativity score.

4.1. Appropriateness, originality and value of ideas across country-language groups

Examining the contribution of appropriateness, originality and value scores to the average student idea score in each country-language group can tell us about cross-cultural differences in relative idea quality. Figure 4.1 shows the composition of the mean idea sum score across tasks in the study in each country-language group, divided into shares that account for the mean appropriateness, originality and value score of the average idea (Box 4.2). On average across country-language groups, close to 40% of the mean idea sum score is derived from idea appropriateness, with the relative shares of value and originality scores contributing a smaller proportion towards the mean idea sum score (32% and 29%, respectively).

These differences in relative contribution are notable given that the range of scores for appropriateness is smaller (0-2 points) than for originality and value (0-3 points), respectively. In real terms, students tend to score higher in appropriateness than they do in originality and value, on average, which suggests that they find it comparatively easier to come up with appropriate ideas than novel or useful ideas (Table B2.2). This is probably as expected, as coming up with an appropriate idea is arguably easier than thinking of an original or useful one. Some learning progressions for creative thinking also typically consider generating appropriate ideas as the first stage of proficiency in creative thinking (Krstic, 2024[25]). These results may also be a function of the tasks included in the study: students are expected to come up with appropriate ideas in all tasks, but not necessarily creative, original or effective ideas in all of them (e.g. 'generate diverse tasks' only require students to think of relevant *different* ideas).

Box 4.2. Which score data are used in this chapter?

The analyses in this chapter draw upon two types of rescored data:

- **Criteria scores:** Each idea within a student response was scored for its appropriateness, originality and value. Appropriateness scores range from 0 to 2 points, whereas originality and value scores range from 0 to 3 points, respectively. A cumulative idea sum score was also produced by aggregating the scores across these three criteria. Tasks classified as 'generate diverse ideas' tasks required students to submit multiple ideas per response, and each idea was scored separately for appropriateness, originality and value.
- **Holistic creativity scores:** Valid student responses were given a score of between 1 (lowest) and 7 points (highest), based on judges' evaluations of the overall creative quality of the response. The response score reflects a relative ranking of creative quality compared to all other responses in that country-language-task group.

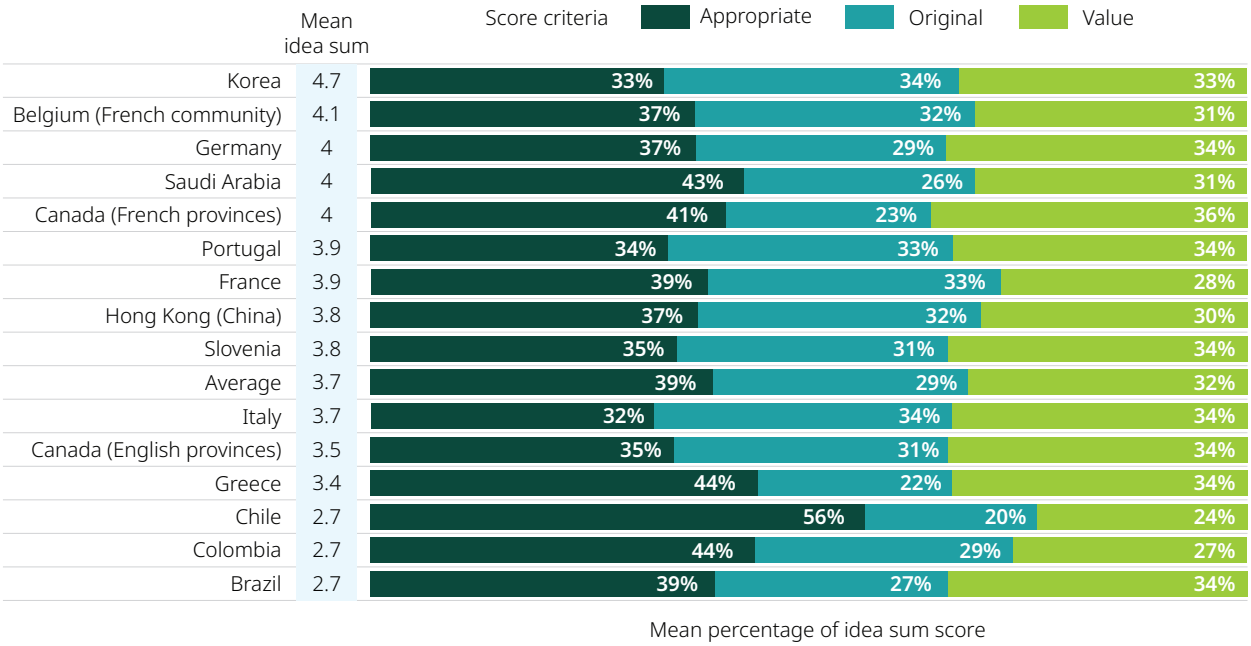
The analyses in this chapter focus on differences in the contribution to and importance of idea appropriateness, originality and value for creative ideas within and across country-language groups. For more information on the scoring methods, see Annex A2.

There are some differences in the general composition of the average idea across tasks across the country-language groups in the study. In Chile, students' appropriateness scores represent a much larger share of the mean idea sum score (56%) compared to other country-language groups (Table B4.1), and in Greece, Colombia and Saudi Arabia, the appropriateness scores represent a relatively larger-than-average share of the mean idea sum score (43-44%). In the case of Colombia and Chile, this might be explained by the overall lower mean idea scores achieved by students compared to those in other country-language groups: while students in these two countries achieve similar mean appropriateness scores to all other students, their mean idea sum score is around 0.9 score points lower than the average idea sum score (Table B2.2). Notably though, students in Brazil scored similarly (and even marginally lower) than students in Colombia and Chile overall, yet appropriateness scores contributed relatively far less to their mean idea sum score. These results imply two things: (1) that students in Chile, Greece, Colombia and Saudi Arabia directed relatively more of their efforts to generating appropriate ideas than original or valuable ones; and/or (2) that judges in these country-language contexts were relatively more generous in their appropriateness scores compared to their scores for originality and value than in other country-language contexts.

On average across country-language groups, there was a modest difference in the relative share of appropriateness and value scores (7 percentage points) (Table B4.1). In many of the country-language groups where appropriateness contributed a relatively large proportion of the mean idea sum score, the relative difference between appropriate and value scores were larger than average – for example in Greece, Saudi Arabia but also France (10-11 percentage points), and particularly in Colombia (17 percentage points) and Chile (32 percentage points). These large differences observed in Chile and Colombia might imply that students found it relatively harder to come up with ideas of added value beyond simply providing an appropriate idea for the task. Alternatively, again, it may be that judges in these country-language groups were relatively less lenient when interpreting the value criterion compared to judges in other country-language groups, or that they incorporated elements in their appropriateness evaluations considered by other judges to better reflect value.

Figure 4.1. Appropriateness scores contribute most to the average idea score, followed by value then originality scores

Mean criteria scores as a relative share of the mean idea sum score (on average across tasks), by country-language group



Notes: Each bar represents 100% of the mean idea sum score on average across tasks in each country-language group. Country-language groups are ordered by mean idea sum score (highest to lowest).

Mean percentages are rounded to the nearest integer.

Source: Table B2.2 and Table B4.1

In most country-language groups, originality scores represented the smallest contribution to students' mean idea sum score (29% on average). This suggests that most students demonstrated a relative weakness in coming up with original ideas compared to appropriate or valuable ideas – especially as some tasks explicitly asked students to focus on idea originality – or it may be that judges were consistently less generous with awarding points for originality compared to the other score criteria across country-language groups. The smallest originality score shares were observed in Chile (20%), Greece (22%), Canada (French provinces, 23%), Saudi Arabia (26%) and Brazil (27%). In only a handful of country-language groups were the relative score shares for originality marginally yet statistically significantly larger than either the appropriateness score share (Italy) or the value score share on average across tasks (Belgium [French community], Colombia, France, Hong Kong [China] and Korea).

4.1.1. Differences by domain context

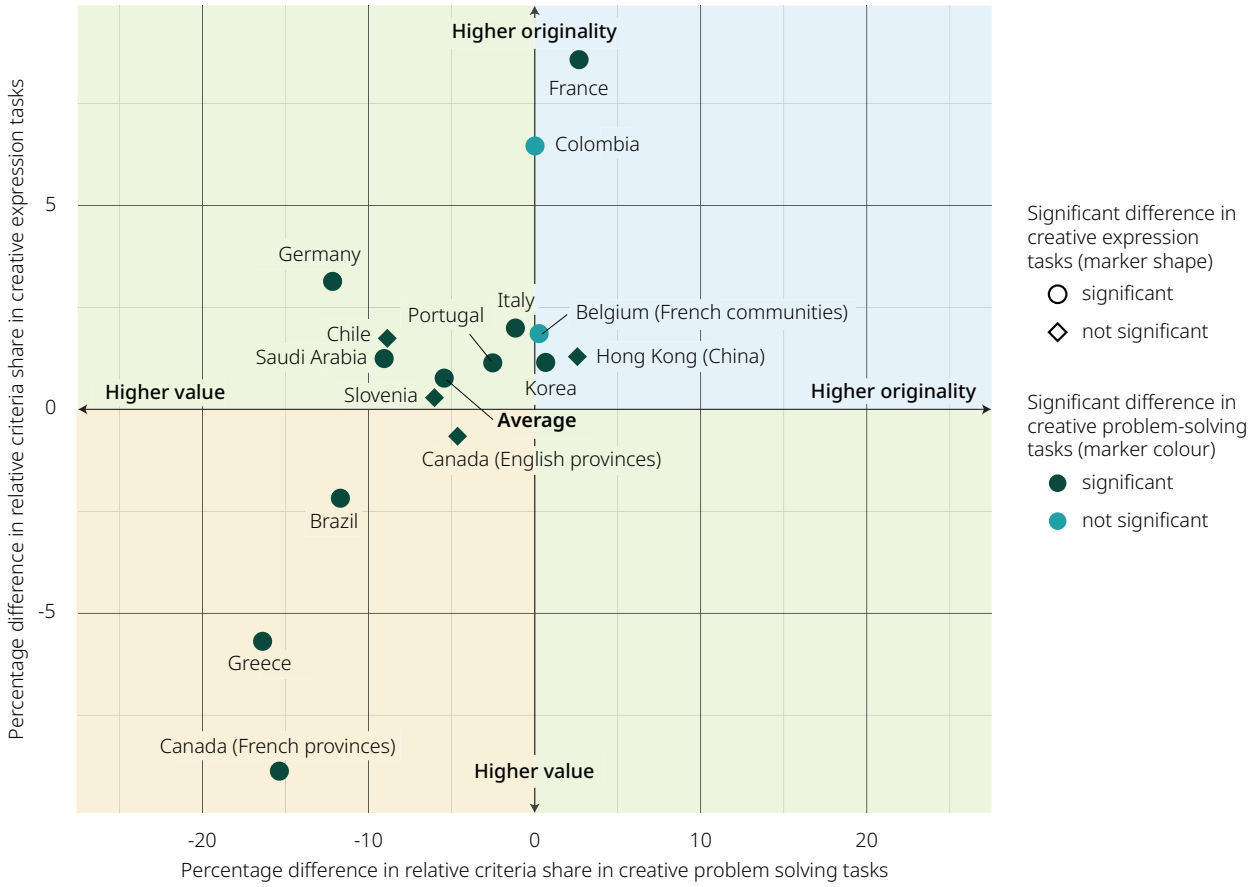
Are there differences in these general patterns when grouping tasks by their domain context? For example, do value scores contribute relatively more to students' overall scores in the problem-solving tasks compared to the creative expression tasks, and vice versa? Do students tend to focus relatively more on originality and relatively less on appropriateness in the expressive tasks? On average across country-language groups, we do observe differences in the relative shares of the criteria scores along these lines, but the differences between task groups are small. In the problem-solving tasks, value contributes only 2 percentage points more to the mean idea sum score than on average across all tasks, at the expense of both appropriateness and originality scores, on average across country-language groups, while in the creative expression tasks, originality scores contribute relatively slightly more and value scores slightly less (both 30%) to the mean idea sum score than on average across all tasks (Table B4.1).

In some country-language groups, these differences by domain context are more pronounced. For example, in Brazil and Saudi Arabia, originality contributes to only 23-24% of the mean idea sum score in problem-solving tasks compared to around 30% of the mean idea sum score in creative expression tasks. An even larger difference in the relative share of originality scores across the two task types is observed in Germany (8 percentage points). In both Colombia and Hong Kong (China), appropriateness scores represent a relatively larger proportion of the average idea score in the creative expression tasks compared to problem-solving tasks (a difference of 8-9 percentage points). And in Chile, Colombia, Germany and Greece, value scores constituted around 7-8 percentage points more of the mean idea sum score in problem-solving tasks than for the expression tasks.

Figure 4.2 shows the difference in the relative shares of originality and value scores, as a proportion of the mean idea sum scores, across tasks in the two broad domain contexts in the PISA CT Rescoring study: creative expression tasks and creative problem-solving tasks. In Canada (English and French provinces), Greece and Brazil, value scores represent a larger share of the total idea sum score than originality scores in both types of task context. Conversely, in France, Korea and Hong Kong (China), originality scores represent a larger share of the idea sum score than value scores in both the expressive and problem-solving task types. In Germany, Saudi Arabia, Chile, Slovenia, Italy and Portugal, as well as on average across country-language groups, originality scores represented a larger share of the mean idea sum score in creative expression tasks than value, but value scores represented a larger share of the mean idea sum score in problem-solving tasks. Interestingly, in no country-language group was the opposite observed.

Figure 4.2. Ideas in creative expression tasks tend to be more original than valuable, whereas ideas in problem solving tasks tend to be the opposite

Mean difference in the relative share of originality and value scores as a proportion of the mean idea sum score, by broad domain grouping and country-language group



Note: Significant differences for creative expression tasks are shown by the shape of the marker (circle/diamond) and for creative problem-solving tasks by the colour of the marker.

Source: Table B4.1

4.2. What makes an idea creative? The relative importance of appropriateness, originality and value within creative ideas across cultural-linguistic contexts

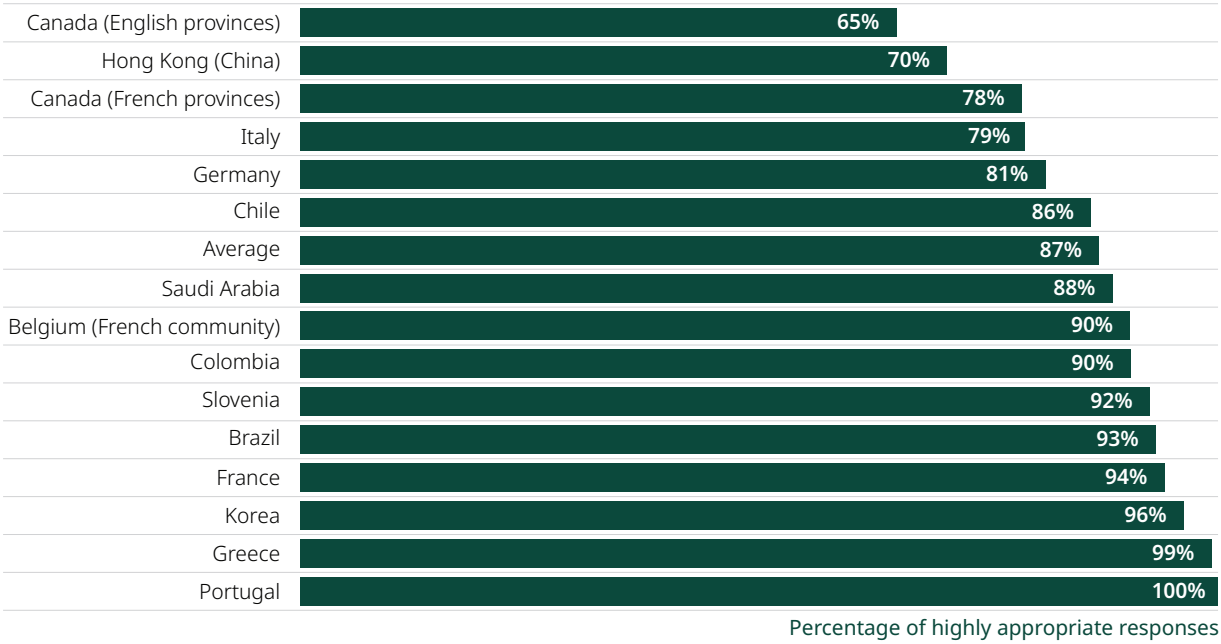
4.2.1. Are creative ideas highly appropriate, original and valuable in all country-language groups?

Chapter 1 explored the characteristics of creative ideas, in general, by examining the interactions between response appropriateness, originality and value scores, with response holistic creativity scores. Our results show that response appropriateness scores are strongly positively associated with judges' holistic creativity scores, on average across tasks and country-language groups (r=0.7), though the strength of this association varies by country-language group – ranging from 0.5 in Canada (French and English provinces) to 0.8 in Colombia, Greece, Korea, Portugal and Slovenia (Table B1.2). Response originality and response value scores were also similarly correlated with response holistic scores (r=0.7), on average, with the strength of these associations ranging from 0.6 to 0.9 by country-language group (Table B1.9 and Table B1.10). These relatively large variations in correlation strength between the different scoring criteria and holistic creativity scores suggest that what matters most in determining creative ideas varies by country-language group.

Figure 4.3 shows the percentage of relatively creative responses – i.e. those that achieved a holistic creativity score of between 5 and 7 score points – that also scored highly for appropriateness (rounded average response idea appropriate score of 2 points), with notable differences across country-language groups. In Portugal, Greece and Korea, above 95% of relatively creative responses were highly appropriate, whereas in Canada (both French and English provinces), Hong Kong (China) and Italy, less than 80% of relatively creative responses were highly appropriate.

Figure 4.3. Most relatively creative responses are highly appropriate, in all country-language groups, though the share varies

Percentage of highly appropriate responses amongst relatively creative responses (on average across tasks), by country-language group



Note: Country-language groups are ordered by percentage of highly appropriate responses amongst relatively creative responses (lowest to highest). Results are computed by task first then averaged across tasks within each country-language group. "Relatively creative responses" are those that achieve a holistic creativity score of 5, 6 or 7. "Highly appropriate" responses are those whose average idea appropriateness score is 2 (rounded). For responses to items classified as 'generate diverse ideas' items, appropriateness scores for all ideas within the response were first averaged to give a single appropriateness score per student response. The averaged appropriateness scores were then rounded to the nearest integer.

Source: Table B4.2

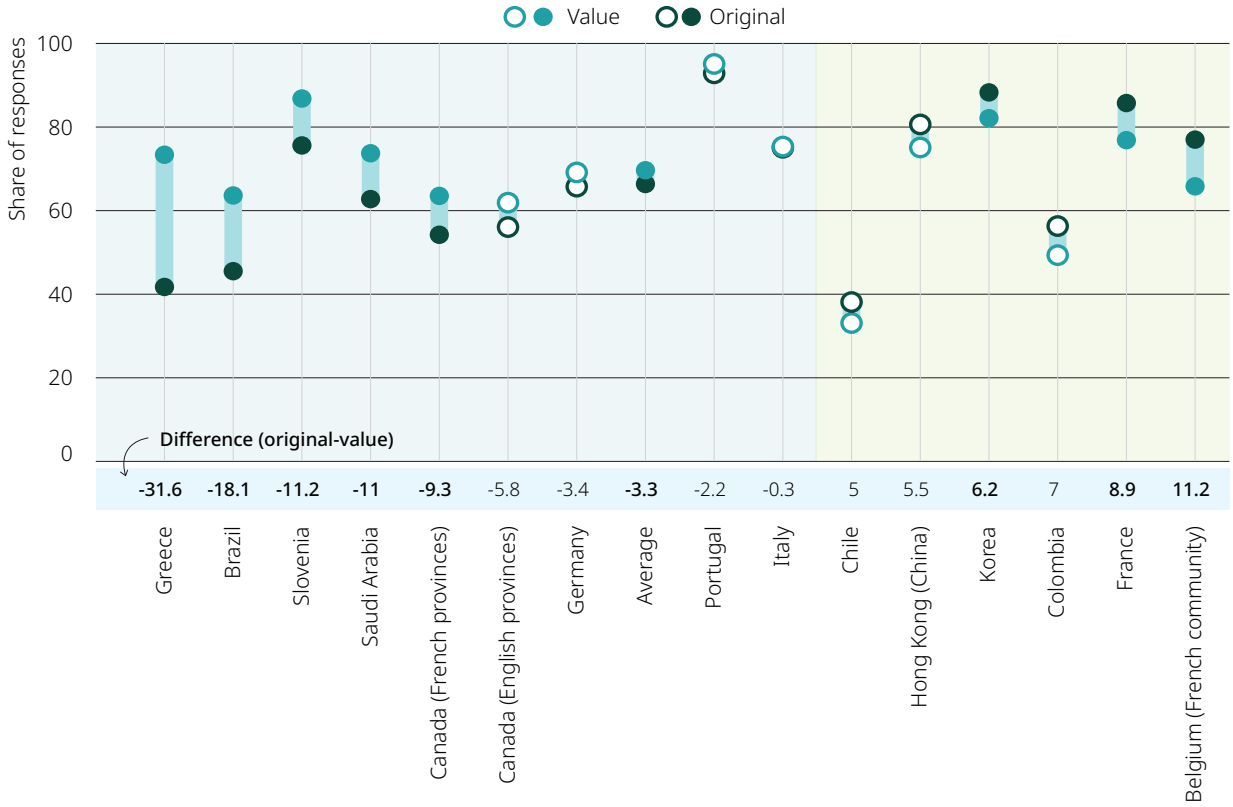
It may be that differences between the holistic and criteria-based scoring methods account for some of these country-language differences, with the quality of ideas in each holistic creativity score category sensitive to the overall performance of students in the country-language group (see Box 4.1). Yet students in all four country-language groups that have less than 80% highly appropriate responses amongst relatively creative responses – Canada (English and French provinces), Hong Kong (China) and Italy – score around or above the average in terms of their mean idea sum scores (Table B2.2). In other words, it is not because students in these countries struggle to think of good ideas that there are less highly appropriate, relatively creative responses. Instead, students and judges in these country-language groups appear more willing to push the boundaries of appropriateness in the pursuit of creativity, or it may simply be that in these country-language groups, appropriateness is not as important as originality and/or value for determining whether a response is creative or not.

Chapter 1 also investigated the relationship between holistic creativity scores and originality and value scores, respectively. The results showed that both criteria are important for defining creative ideas, in general, and that responses at the very highest holistic creativity scores (scores 6 and 7) tended to have a higher mean originality score than value score, on average across tasks and country-language groups. Are there differences in these general observations across country-language groups? Figure 4.4 shows the difference in the shares of highly original and highly valuable responses (i.e. those with a mean response idea originality or value score of at least 2 points, respectively) within relatively creative responses, by country-language group. In around half of the country-language groups included in the study, the shares of highly original and highly valuable responses amongst relatively creative responses are similar and not statistically significantly different – although on average, there is a small statistically significant difference in favour of highly valuable responses amongst relatively creative responses (difference of 3 percentage points). This result may be explained by the larger number of student responses in holistic score category 5 compared to score categories 6 and 7, with responses in score category 5 scoring higher in value than originality on average across country-language groups (Table B1.7 and Table B1.8). These results generally support the idea that both criteria are important for defining creative ideas across socio-cultural contexts.

Two country-language groups showed a much larger difference in the relative shares of highly valuable and highly original responses amongst responses that achieved a holistic creativity score of at least 5: Greece (difference of 32 percentage points) and Brazil (difference of 20 percentage points). In these two countries, there were far more highly valuable responses amongst relatively creative responses than highly original ones. This is likely because students in both Greece and Brazil struggled to come up with original ideas in general, with the mean idea originality score in both countries around 0.3 score points lower than on average across country-language groups. In contrast, their mean value scores were around (Greece) or only slightly below (Brazil) the country-language group average (Table B2.2). Canada (French provinces) also recorded a statistically significant larger share of highly valuable responses amongst relatively creative responses (difference of 10 percentage points). Only in Belgium (French community), France and Korea were there significantly more highly original responses than highly valuable ones amongst student responses that achieved a score of at least 5 points on the holistic creativity scale.

Figure 4.4. Relatively creative responses tend to be more highly valuable than original, though not in all country-language groups

Difference in the relative share of highly original/valuable responses amongst relatively creative responses (on average across tasks), by country-language group



Note: Country-language groups are ordered by difference in relative share of highly original/valuable responses (original-value) amongst relatively creative responses. Values that are statistically significant are indicated by a filled marker. Results are computed by task first then averaged across tasks within each country-language group. "Relatively creative responses" are those that achieve a holistic creativity score of 5, 6 or 7. For 'generate diverse ideas' items that contain multiple ideas per response, the criteria (originality/value) scores for all ideas were averaged to give a single criteria score per student response. "Highly original/valuable" responses are those with an average response criteria score of at least 2 score points.

Source: Table B4.2

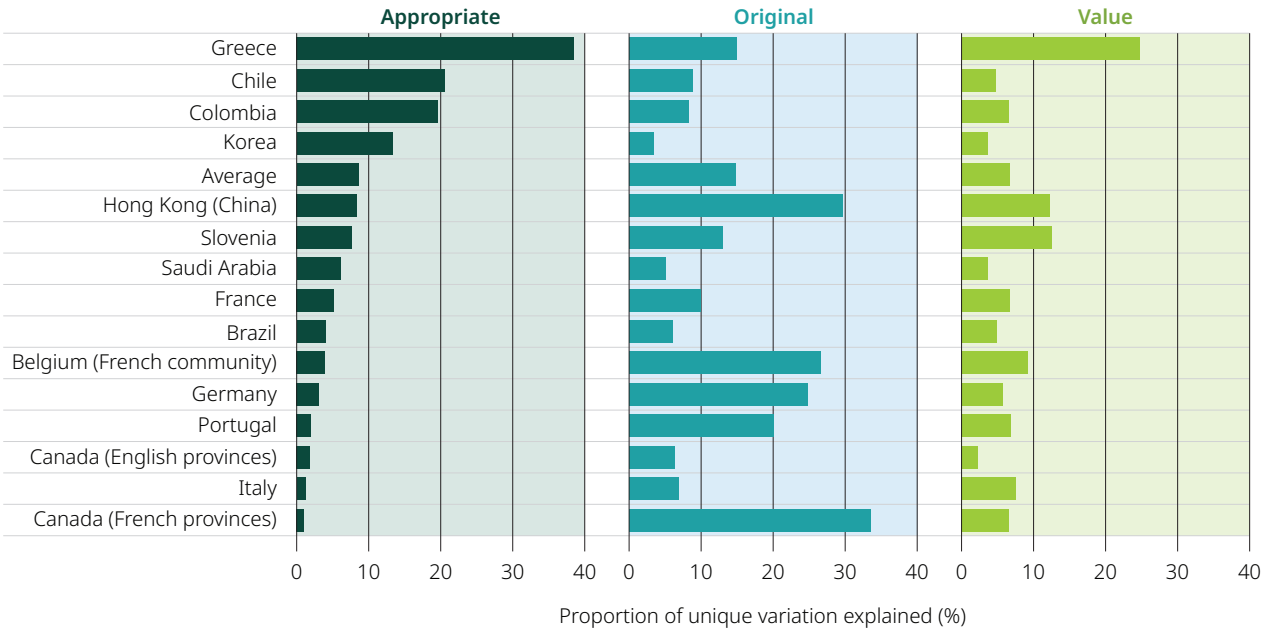
4.2.2. Variation in holistic creativity scores explained by appropriateness, originality and value across country-language groups

Given that the findings shown in Figure 4.3 and Figure 4.4 are sensitive to the overall success of students in each country-language group, another way of examining the relative importance of appropriateness, originality and value in defining creative responses overall is to examine and compare the holistic scores of students with similar criteria scores. Figure 4.5 shows the unique variation in judges' holistic creativity scores explained by appropriateness, originality and value, respectively, in each country-language group. It shows the additional variation in holistic creativity scores explained by students' response appropriateness scores, for example, when added to a linear regression model that already includes originality and value scores.

Overall, the results show that there is significant variation across country-language groups in the extent to which response appropriateness, originality and value scores explain holistic creativity scores. These results also do not seem to be highly linked to student success or judge leniency in the scoring. On average, the largest proportion of unique variation in holistic creativity scores is explained by originality scores (around 15%), followed by appropriateness (around 9%) and then value (around 7%). In Canada (French provinces) originality scores explain well over 30% of the unique variation in holistic creativity scores, and in Hong Kong (China), Belgium (French community) and Germany, originality scores explain between 25-30%. In contrast, in Korea and Saudi Arabia, originality scores explain only 5% or less additional variation in holistic creativity scores in a model that already includes response appropriateness and value scores.

Figure 4.5. The relationship between holistic creativity scores and response appropriateness, originality and value scores varies significantly by country-language group

Proportion of the unique variation in holistic creativity scores explained by mean response appropriate, original and value scores (all tasks), by country-language group



Note: Country-language groups are ordered by the proportion of unique variation in holistic creativity scores explained by appropriate scores (from highest to lowest). The proportion of unique variation in holistic creativity score describes the proportion of variance in holistic creativity score additionally explained by a given score criterion beyond the variation explained by the other score criteria. These results are likely attenuated given that judges were also instructed to consider idea flexibility when assigning holistic creativity scores to responses to 'generate diverse ideas' (GDI) items. Results were computed on the pooled responses across all tasks in each country-language group. For GDI items that contain multiple ideas per response, the criteria (appropriate/originality/value) scores for all ideas were first averaged to give a single criteria score per student response.

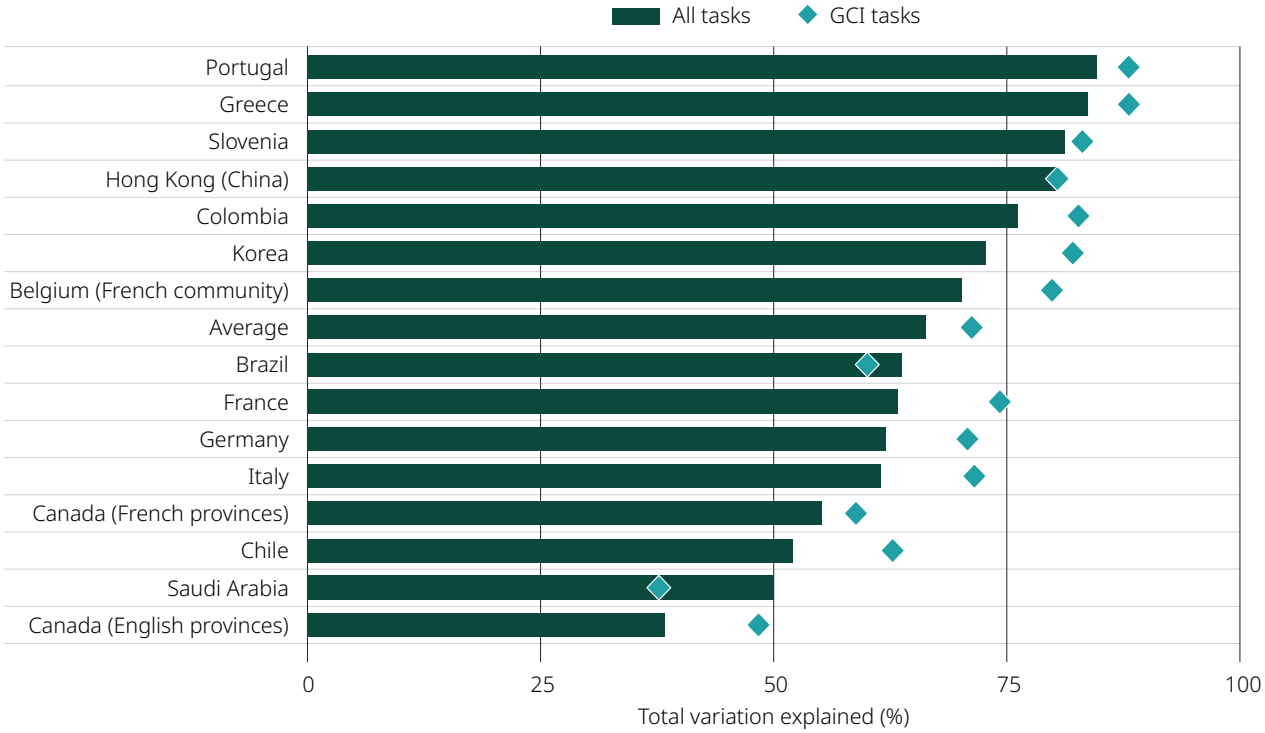
Source: Table B4.3

The importance of appropriateness scores in explaining holistic creativity scores appears to be the most variable score criteria across the country-language groups included in the PISA CT Rescoring study, with response appropriateness explaining the largest relative proportion of unique variation in holistic creativity scores in Greece (39%), Chile (21%), Colombia (20%) and Korea (13%) yet only marginal amounts (less than 2%) in Italy and Canada (French and English provinces). Conversely, value scores seem to explain a relatively modest but consistent proportion of the unique variation in holistic creativity scores across country-language groups. Only in Greece do value scores seem to explain a relatively large proportion of the unique variation in holistic creativity scores (around 25%) (Table B4.3).

Country-language groups also vary in the extent to which appropriateness, originality and value scores collectively explain the total variation in holistic creativity scores. As shown in Figure 4.6, in Portugal, Greece, Slovenia, Hong Kong (China) and Colombia, over 75% of the total variation in holistic creativity scores can be explained by judges' response criteria scores, while less than half of the variation in holistic creativity scores in Canada (English provinces) can be explained by these three criteria scores. One additional criteria that would clearly influence holistic creativity scores for responses to some items ('generate diverse ideas' items) is idea flexibility, given that responses that contained very different ideas may have achieved a higher holistic score and judges were given guidelines regarding the maximum holistic score to award in cases where some ideas were missing from a response (see Annex A2). However, even when looking at the total variation explained in 'generate creative ideas' tasks only (see markers in Figure 4.6), there remains unexplained variation in students' holistic creativity scores across all country-language groups.

Figure 4.6. Appropriateness, originality and value explain most – but not all – of the variation in holistic creativity scores

Total variation in response holistic creativity score explained by mean response appropriate, original and value scores (all tasks and 'generate creative ideas' tasks only), by country-language group



Note: Country-language groups are ordered by proportion of total variation explained across all tasks (highest to lowest). Bars show the total variation in holistic creativity scores explained by criteria (appropriate, originality and value) scores across all tasks (pooled responses), blue markers show the total variation explained in holistic creativity scores by these factors across 'generate creative ideas' tasks only (pooled responses). For 'generate diverse ideas' items that contain multiple ideas per response, the criteria scores for all ideas were first averaged to give a single criteria score per student response.

Source: Table B1.18 and Table B4.4

Once again, as discussed in Chapter 1 of this report, these results point towards the influence of other important factors that are not captured by the score criteria of appropriateness, originality and value in determining the overall creativity of a student response – for example, the inclusion of domain-relevant elements within a response, or other response characteristics that might implicitly or explicitly influence judges' holistic creativity evaluations (e.g. sophistication of vocabulary, response length, etc.).

4.3. Measuring creativity across cultures: a final word

Beyond the specific findings of the PISA CT Rescoring project, the results in this chapter highlight the complexity of measuring creativity – both in general, and specifically in a cross-culturally valid and reliable way. Although coding teams in our study completed small anchor training exercises and were required to achieve a certain level of within-country inter-rater reliability in their coding (see Annex A2), it is not easy to disentangle whether the cross-cultural differences observed in this chapter stem from differences in the preferences and priorities of students or differences in how judges understand and implement the evaluation criteria. Prior researchers have noted not only differences in performance expectations across cultures more broadly when it comes to creative work (Shao et al., 2019[5]; Lubart, 1998[4]; Rudowicz, 2003[6]; Ng, 2003[8]; Niu and Sternberg, 2003[2]), but the effects of socio-cultural rater bias in creativity assessments (Kaufman et al., 2010[24]; McCarthy, Chen and McNamee, 2018[26]).

The country-language groups included in the PISA CT Rescoring study represent a diverse set of socio-linguistic and cultural groups across North America, Latin America, Europe, the Middle East and South-East Asia. Despite efforts to reduce cross-cultural bias in the design of the PISA tasks throughout the test development process (see Foster and Piacentini, 2025[27]; OECD, 2024[28] for more), task and content familiarity will influence both students' capacity to generate creative ideas and judges' capacity to score students' ideas in a cross-culturally consistent way. In other words, both familiarity with the test materials and culture-related task misinterpretations resulting from different norms or social conventions will influence the accuracy of creativity assessments and the comparability of their results across student groups (Shao et al., 2019[5]). Designing and administering culturally appropriate tasks is therefore important for addressing performance bias across student groups and improving the validity of assessment results. In the context of education specifically, this would mean developing tasks that are more clearly connected to the curriculum and typical classroom experiences of different student populations.

The scoring method used in the PISA 2022 Creative Thinking assessment was developed in such a way to optimise the comparability of the results across diverse socio-cultural and linguistic contexts and to facilitate the implementation of a large-scale, human-scoring operation in 64 countries and economies. This was namely achieved by using statistical infrequency as a more objective proxy for originality, by not including any evaluation of response value, and by developing a relatively simple, atomised scoring procedure with a limited number of score codes. The complex trade-offs resulting from these design choices, including the granularity, sensitivity and interpretation of the PISA scores, are discussed in detail elsewhere (Foster and Piacentini, forthcoming[29]), but the PISA scoring method resulted in a sufficiently reliable approach to scoring and scaling the PISA data.

The scoring methods used in the PISA CT Rescoring study differ significantly by design (see Box 4.1) and in how they either allow for or mitigate potential cross-cultural differences in creativity. The holistic scoring method attempts to rank student responses within a country-language-task group based on a variety of factors, deemed important by each country-language coding team, that contribute to a response's overall creative quality. This ranking exercise, and the relative weighting of the factors influencing such a ranking, are inherently socio-culturally variable and intended to capture and reflect the cultural specificity of creativity in each context. Therefore, in the holistic scoring method, differences in widespread knowledge and social norms mean that responses that are considered creative in one country-language context may not be considered creative in another (Shao et al., 2019[5]). In contrast, the criteria-based method was developed as a more analytic and standardised approach to scoring creative ideas across country-language teams, but that nonetheless addressed one of the major omissions of the PISA scoring approach (the lack of value criteria). The OECD developed standardised guidelines on how to interpret the appropriateness, originality and value criteria for each task included in the study, given the form and response type varied from task to task. The aim of these guidelines was to establish a shared understanding of the respective scoring criteria for each task and promote their consistent application, both within and across coding teams. While the analyses in this chapter focus on understanding the relationships between the two methods – precisely to understand how the core criteria of appropriateness, originality and value interact with judges' holistic evaluations – choosing an appropriate scoring method for any measurement context is ultimately a question of the assessment goals, and the extent to which cross-cultural differences in how creativity is understood and manifests need to be mitigated or accommodated.

Key findings and implications

- The criteria of appropriateness, originality and value remain useful and relevant for evaluating creative ideas across diverse country-language groups, but their relative importance may differ across socio-cultural contexts.
  - Cross-cultural differences in the interpretation of appropriateness, originality and value might stem from different cultural expectations of judges or different preferences of student groups.
- Students find it relatively easier to come up with appropriate ideas than original or valuable ideas – with appropriateness representing 39% of students' mean idea sum scores, on average. However, the proportion of highly appropriate, relatively creative responses varied significantly across country-language groups – from 65% to 100% of creative ideas.
  - The extent to which appropriateness defines creative ideas may vary significantly amongst students and judges from different socio-linguistic and cultural backgrounds. The appropriateness of creative ideas in each country-language group does not seem strongly linked to the overall mean scores of students.
- Originality explained around 15% (the largest share) of the unique variation in holistic creativity scores, on average. However, students demonstrate a relative weakness in coming up with original ideas, which was particularly evident in some countries – possibly because judges are more demanding in their evaluations of what it means to be original in such tasks.
  - Both students and those evaluating their work (e.g. educators) might benefit from support in understanding how originality can be expressed in appropriate, task-relevant ways. In an educational context, Bahar and Maker (2025[30]) suggest linking originality to uncommon ways of demonstrating specific learning objectives in the curriculum, rather than broader norm-based expectations of originality.
- In a measurement context, socio-cultural norms and expectations matter: tasks must be culturally relevant, and evaluation criteria should be contextualised in culturally appropriate ways.
  - Contextualised scoring criteria should also be clearly communicated to students and judges, so that both the creators and evaluators of creative work share an understanding of what constitutes a culturally-relevant, creative idea for a given context.

References

Amabile, T. (2012), *Componential theory of creativity*, No. 12-096, Harvard Business School, http://www.hbs.edu/faculty/Publication%20Files/12-096.pdf. [10]

Amabile, T. (1983), "The social psychology of creativity: A componential conceptualization", *Journal of Personality and Social Psychology,* Vol. 45/2, pp. 357-376, https://doi.org/10.1037/0022-3514.45.2.357. [9]

Amabile, T. and M. Pratt (2016), *The dynamic componential model of creativity and innovation in organizations: Making progress, making meaning*, https://doi.org/10.1016/j.riob.2016.10.001. [11]

Bahar, K. and C. Maker (2025), "A new perspective on scoring children's originality: a standards-based criterion-referenced assessment approach", *Frontiers in Psychology,* Vol. 16, https://doi.org/10.3389/fpsyg.2025.1545396. [30]

Beghetto, R. and J. Kaufman (2014), "Classroom contexts for creativity", *High Ability Studies,* Vol. 25/1, pp. 53-69, https://doi.org/10.1080/13598139.2014.905247. [18]

Beghetto, R. and J. Kaufman (eds.) (2010), *How to discourage creative thinking in the classroom,* Cambridge University Press, https://doi.org/10.1017/CBO9780511781629.002. [19]

Beghetto, R. and J. Kaufman (2010), *Nurturing creativity in the classroom,* Cambridge University Press. [17]

DeCoker, G. (2000), "Looking at U.S. education through the eyes of Japanese teachers", *Phi Delta Kappan,* Vol. 81. [16]

Foster, N. and M. Piacentini (2025), "Assessing creative thinking at scale and across culture", in Katz-Buonincontro, J. and T. Kettler (eds.), *Oxford Handbook of Creativity and Education,* Oxford Handbooks, Oxford University Press. [27]

Foster, N. and M. Piacentini (forthcoming), "PISA 2022 creative thinking: What it is, what it is not, and what next?", *The Journal of Creative Behavior*, https://doi.org/10.1002/jocb.70068. [29]

Kaufman, J., J. Baer and J. Cole (2009), "Expertise, domains, and the consensual assessment technique", *Journal of Creative Behavior*, Vol. 43/4, pp. 223-233, https://doi.org/10.1002/j.2162-6057.2009.tb01316.x. [21]

Kaufman, J. et al. (2008), "A comparison of expert and nonexpert raters using the consensual assessment technique", *Creativity Research Journal,* Vol. 20/2, pp. 171-178, https://doi.org/10.1080/10400410802059929. [20]

Kaufman, J. et al. (2010), "In the eye of the beholder: Differences across ethnicity and gender in evaluating creative work", *Journal of Applied Social Psychology,* Vol. 40/2, pp. 496-511, https://doi.org/10.1111/j.1559-1816.2009.00584.x. [24]

Krstic, S. (2024), *Putting creative thinking at the core of the English school curriculum: An exploratory study.,* Australian Council for Educational Research, https://research.acer.edu.au/cgi/viewcontent.cgi?article=1077&context=ar_misc. [25]

Lloyd-Cox, J., A. Pickering and J. Bhattcharya (2022), "Evaluating creativity: How idea context and rater personality affect considerations of novelty and usefulness", *Creativity Research Journal,* Vol. 34/4, pp. 373-390. [23]

Long, H. (2014), "More than appropriateness and novelty: Judges' criteria of assessing creative products in science tasks", *Thinking Skills and Creativity,* Vol. 13, pp. 183-194, https://doi.org/10.1016/j.tsc.2014.05.002. [22]

McCarthy, M., C. Chen and R. McNamee (2018), "Novelty and usefulness trade-off: Cultural cognitive differences and creative idea evaluation", *Journal of Cross-Cultural Psychology,* Vol. 49/2, pp. 171-198, https://doi.org/10.1177/0022022116680479. [26]

Ng, A. (2003), "A cultural model of creative and conforming behavior", *Creativity Research Journal*, Vol. 15/2&3, pp. 223-233, https://doi.org/10.1080/10400419.2003.9651414. [8]

Niu, W. and R. Sternberg (2003), "Societal and school influences on student creativity: the case of China", *Psychology in the Schools*, Vol. 40/1, pp. 103-114, https://doi.org/10.1002/pits.10072. [2]

OECD (2024), *PISA 2022 Technical Report,* PISA, OECD Publishing, https://doi.org/10.1787/01820d6d-en. [28]

Plucker, J., R. Beghetto and G. Dow (2004), "Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research", *Educational Psychologist,* Vol. 39/2, pp. 83-96. [1]

Rinne, T., G. Steel and J. Fairweather (2013), "The role of Hofstede's individualism in national-level creativity", *Creativity Research Journal,* Vol. 25/1, pp. 126-136, https://doi.org/10.1080/10400419.2013.752293. [7]

Rudowicz, E. (2003), "Creativity and culture: A two-way interaction", *Scandinavian Journal of Educational Research*, Vol. 47/3, pp. 273-290, https://doi.org/10.1080/00313830308602. [6]

Shao, Y. et al. (2019), "How does culture shape creativity? A mini-review", *Frontiers of Psychology*, Vol. 10, https://doi.org/doi: 10.3389/fpsyg.2019.01219. [5]

Sternberg, R. (2006), "The nature of creativity", *Creativity Research Journal.*, Vol. 18/1, pp. 87-98, https://www.cc.gatech.edu/classes/AY2013/cs7601_spring/papers/Sternberg_Nature-of-creativity.pdf. [14]

Sternberg, R. (ed.) (1998), *Creativity across cultures,* Cambridge University Press, https://doi.org/10.1017/CBO9780511807916.019. [4]

Sternberg, R. and T. Lubart (1995), *Defying the crowd: Cultivating creativity in a culture of conformity,* Free Press. [13]

Sternberg, R. and T. Lubart (1991), "An investment theory of creativity and its development", *Human Development*, Vol. 34/1, pp. 1-31, https://doi.org/10.1159/000277029. [12]

Wong, R. and W. Niu (2013), "Cultural difference in stereotype perceptions and performances in nonverbal deductive reasoning and creativity", *The Journal of Creative Behavior,* Vol. 47, pp. 41-59, https://doi.org/10.1002/jocb.22. [3]

Zhou, J. and Y. Su (2010), "A missing piece of the puzzle: The organizational context in cultural patterns of creativity", *Management and Organization Review,* Vol. 6/3, pp. 391-413, https://doi.org/10.1111/j.1740-8784.2010.00192.x. [15]

05

Thinking outside or inside the box? Patterns in idea flexibility and fixations in problem-solving tasks

In the context of measuring creativity and creative thinking, ideational fluency (i.e. being able to think of many ideas) and ideational flexibility (i.e. being able to think of very different ideas) are often used as proxies of creative abilities in divergent thinking tasks. Chapter 2 of this report focused on unpacking differences in the creative success of students in divergent and convergent thinking tasks included in the PISA CT Rescoring study, finding that students were more likely to propose creative ideas in the divergent thinking tasks in the study, on average across country-language. While students may ultimately have been more successful at thinking *creatively* in the divergent thinking tasks in the study, over a third of students across country-language groups were not able to come up with sufficiently different ideas according to judges in these tasks (two or three, depending on the task requirements) (Table B5.1).

One of the reasons individuals may find it difficult to think flexibly stems from a phenomenon known as "fixation effect", which refers to the idea generation process being overly constrained to existing or obvious "known" solutions and thus limiting the generation of new and alternative solutions (Smith, 1995[1]). Fixation effects have been well documented in research on design thinking (Purcell and Gero, 1996[2]; Jansson and Smith, 1991[3]; Viswanathan and Linsey, 2011[4]), and research in cognitive psychology has also identified how various factors might induce fixation effects in creative tasks. These include through functional priming (Adamson, 1952[5]; Defeyter and German, 2003[6]) or activating individuals' knowledge and memory by exposing them to existing solution examples or others' ideas (Smith, Ward and Schumacher, 1993[7]; Agogué, 2015[8]). Studies have demonstrated that the type of examples shown to individuals can serve to strengthen or mitigate fixation effects (Agogué et al., 2014[9]), and that the age and expertise of individuals may determine how these factors interact with fixation effects (Cassotti et al., 2016[10]; Defeyter and German, 2003[6]; Viswanathan and Linsey, 2011[4]).

While the judges involved in the PISA CT Rescoring project awarded student responses a score for ideational flexibility as part of the criteria-based scoring method (see Annex A2 for more information), evaluating whether ideas are 'sufficiently' different or not in a cross-culturally valid way is challenging without a comprehensive rubric. Several of the studies cited above that focus on examining fixation effects employ a Concept-Knowledge (C-K) schema (Hatchuel and Weil, 2009[11]) to investigate the extent to which individuals can think flexibly and avoid idea fixations in divergent thinking tasks. The C-K approach involves creating an exhaustive mapping of combinations of concepts and knowledge within a given problem and solution space (see Box 5.1). This objective mapping, often conceptualised as a C-K 'tree' with branches, sub-branches and nodes, allows judges to map ideas onto a well-defined schema from which solution pathways can be inferred, as well as more granular idea flexibility and originality metrics.

We applied the C-K approach to two of the divergent thinking, problem-solving tasks included in the PISA CT Rescoring study: one relatively familiar social problem-solving task (Task 5 – *Food Waste*) and one more difficult, engineering-type task (Task 7 – *The Exhibit*). To what extent can students really 'think outside of the box' in different problem-solving tasks? Are different student groups more susceptible to fixation effects? Do we observe strong socio-linguistic differences in the types of ideas that students think of, or are students' ideas more similar than we might expect across diverse socio-cultural and linguistic contexts?

5.1. How diverse are students' ideas in problem-solving tasks?

5.1.1. Food Waste: a social problem-solving task in a familiar context

Task 5 in the PISA CT Rescoring study – *Food Waste* – was classified as a social problem-solving task in the PISA 2022 Creative Thinking assessment (Figure 5.1). The task presents students with the issue of food wastage in supermarkets and asks them to propose three different ways to address this problem. The problem context is relatively familiar and accessible: many if not all students that sat the PISA 2022 test would be familiar with supermarkets or grocers' shops that sell fresh food items, and citizens including young adults in many countries have become increasingly socially engaged with issues related to reducing waste, recycling, etc. over the past decade. Students were generally successful at thinking of "sufficiently different" ideas for this task, according to judges' criteria-based scores (see Box 5.1), with 65% of students on average across country-language groups achieving the maximum idea flexibility score and just 14% of all students unable to provide at least two different ideas (Table B5.2).

Data scored using the C-K Tree approach paint a much more nuanced story of the diversity of students' ideas, both within and across country-language groups. The full C-K tree schema for *Food Waste* contained 8 primary solution paths (i.e. 'branches'), with 25 sub-branches and 114 pre-defined end nodes in total (see Box 5.1 and Annexes C1 and C2). Of these 25 conceptually different sub-branches, just under 43% of student ideas on average across country-language groups were categorised within the same, single sub-branch: "Sub-branch E – distribute unsold produce for free" (Figure 5.2). In Chile and Portugal, over half of all ideas proposed by students corresponded to this sub-branch (Table B5.3). In some ways, the relatively high percentage of ideas within a single sub-branch is understandable, given that distributing unsold produce for free represents one of the most obvious solution paths for addressing the issue of food waste, and given that students were instructed to think of *different* ideas not necessarily *creative* ones. It therefore makes sense that many students chose an obvious solution for at least one of their three ideas.

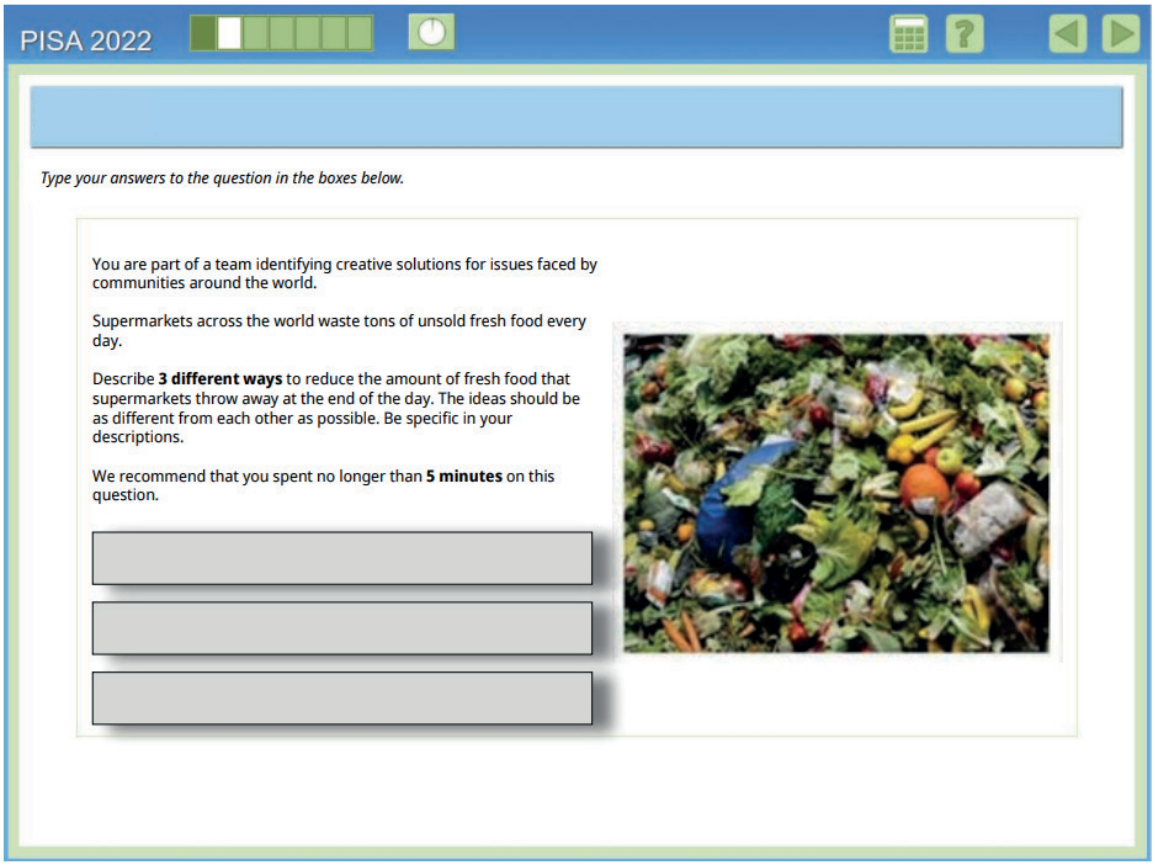
Box 5.1. Which score data are used in this chapter?

The analyses in this chapter draw upon two types of rescored data from the PISA CT Rescoring study:

- Conceptual-Knowledge (C-K) Tree codes: These data are categorical and describe, for each idea within a student response, its corresponding sub-branch and node on the task-specific Conceptual-Knowledge tree. The process for developing the C-K trees for the two tasks discussed in this chapter is described in Annex A2, and the full C-K tree schema and accompanying list of nodes for both tasks are detailed in Annexes C1 to C4.
- Criteria-based scores: Some analyses draw upon response flexibility scores attributed by judges based on the number of "sufficiently different" ideas provided by the student within a response.

This chapter focuses on examining ideational flexibility and identifying fixation paths in student responses to two problem-solving tasks across country-language groups.

Figure 5.1. Food Waste task

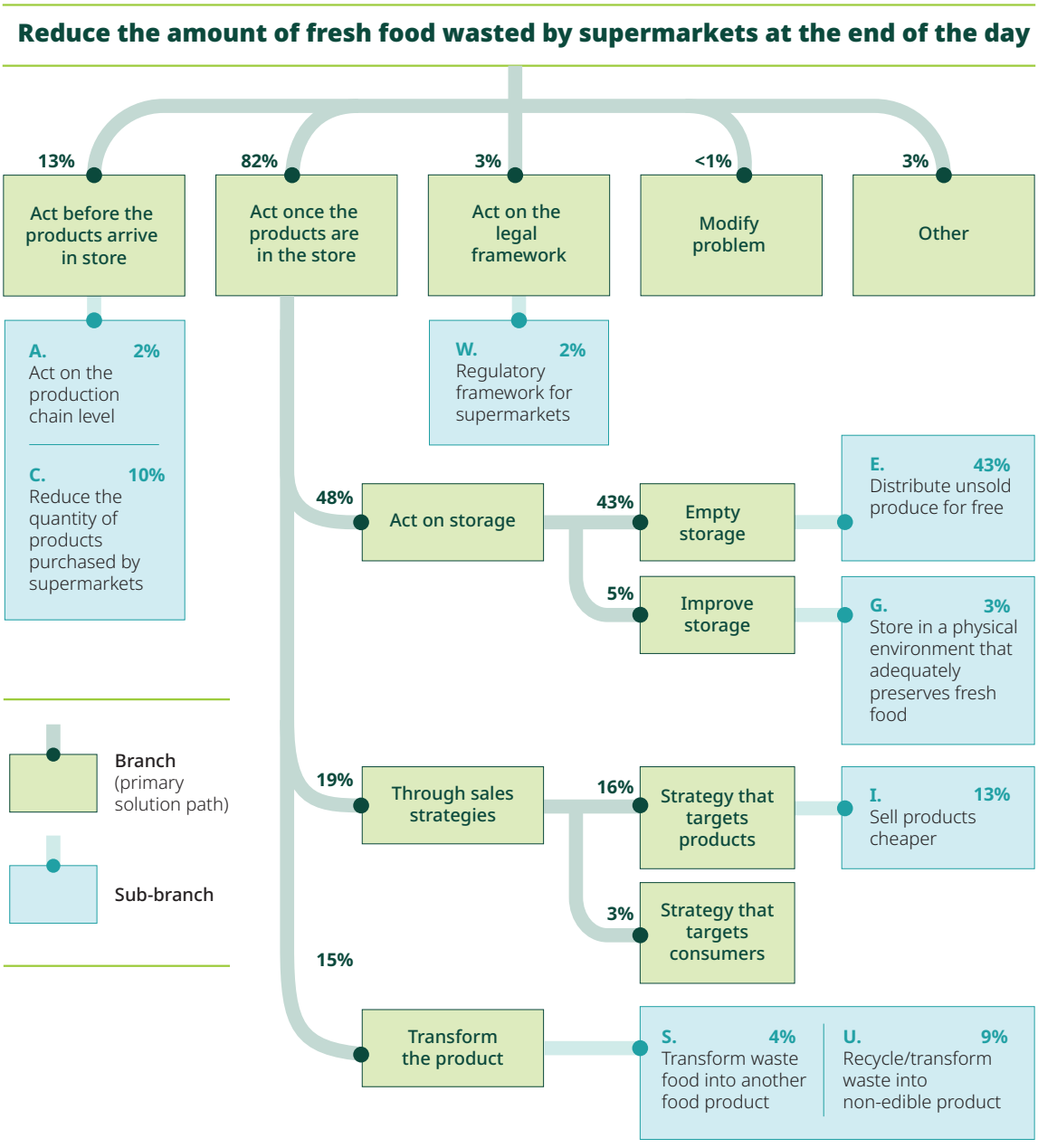


Source: PISA 2022 Creative Thinking test.

Perhaps more surprisingly than the large share of ideas across country-language groups within the same sub-branch is that, even amongst all ideas within that sub-branch, students were remarkably consistent in their ideation: around 4 in 10 ideas on average across all country-language groups corresponded to the same specific node of the C-K Tree ("give produce to people directly for free") (Table B5.5). Only in Canada (English provinces) and Saudi Arabia did students think of slightly different ideas within that same sub-branch most frequently ("donate produce to charities").

Figure 5.2. Nearly half of all student ideas in *Food Waste* fall within the same sub-branch of the C-K tree

Percentage of student ideas within each branch and select sub-branches, on average across country-language groups



Note: Only sub-branches with over 2% of all student ideas on average across country-language groups are shown. Percentages shown in the figure are rounded to the nearest integer. The full C-K Tree schema for this task is detailed in Annex C1.

Source: Table B5.3

Table 5.1 summarises the 5 most common idea nodes across the full C-K schema across country-language groups in the study. Overall, of the top 5 most common ideas (i.e. nodes), three were situated within sub-branch E: "give produce directly to people for free" (node ee; 23% ideas on average), "donate produce to charities" (node ea; 12% ideas on average) and "donate produce to animals or animal/agricultural structures" (node ei; 7% ideas on average). Despite the C-K Tree schema defining 114 end nodes *a priori*, the 5 most common nodes account for 62% of all ideas. However, only the 3 most common nodes on average across country-language groups account for more than 10% of all ideas each, with the rest of students' ideas spread across a more diverse range of nodes.

Table 5.1. Three out of the five most common ideas in *Food Waste* fall within the same sub-branch

Percentage of all student ideas within the 5 most common nodes (on average across country-language groups) on the C-K Tree for *Food Waste*

Node description	%	Sub-branch (node)
Give produce to people directly for free	23%	E (ee)
Lower the price of individual products	13%	I (ia)
Donate to charities	12%	E (ea)
Donate to animals or agricultural/animal structures	7%	E (ei)
Supermarkets to buy fewer products	7%	C (ca)

Note: Percentages are rounded to the nearest integer. For a full description of all nodes for this task, see Annex C2.
Source: Table B5.6

Beyond the sub-branch "distribute unsold produce for free", only two other sub-branches amongst all those available in the problem space contain more than 10% of all student ideas each, on average across country-language groups. These are "sell products cheaper" (sub-branch I) and "reduce the quantity of products purchased by supermarkets" (sub-branch C), respectively. Table 5.2 describes the 5 sub-branches with the largest share of ideas, on average across country-language groups.

Table 5.2. Over two thirds of all ideas focus on addressing the problem of food waste once products are already in the supermarket

Percentage of all student ideas within the 5 most common sub-branches (on average across country-language groups) on the C-K Tree for *Food Waste*

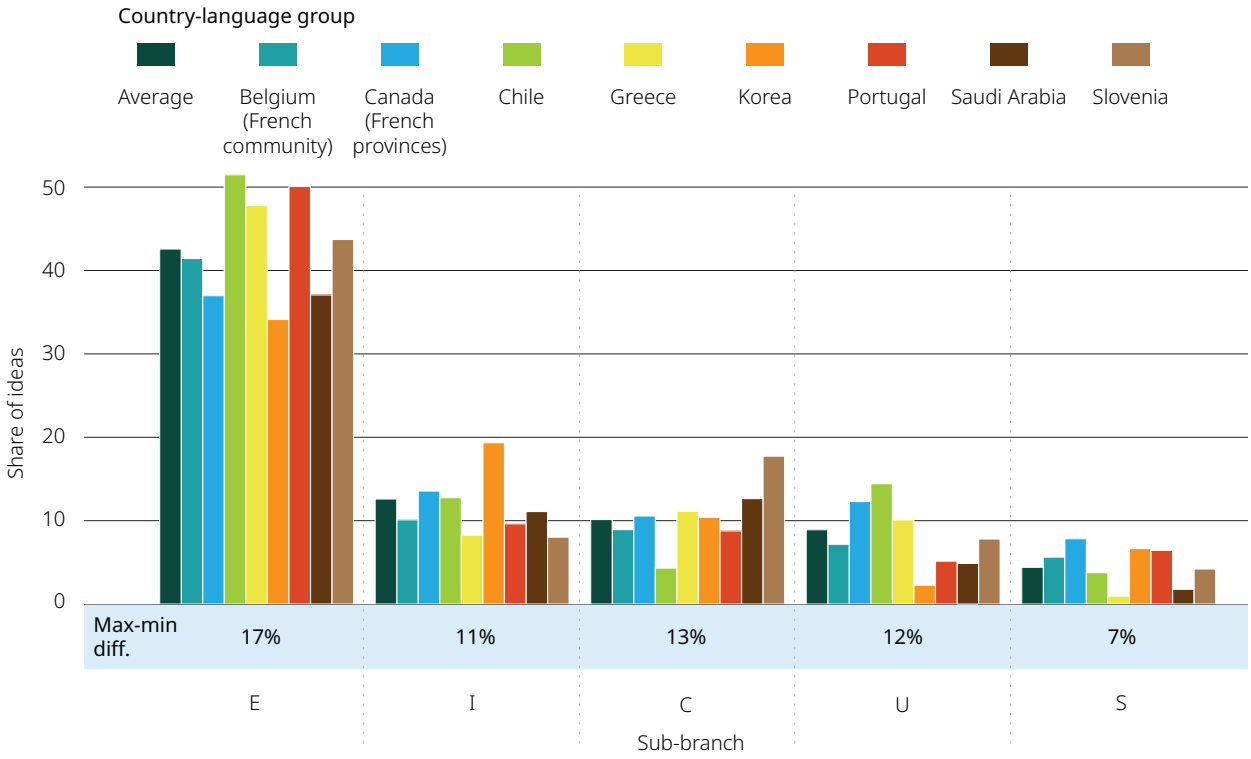
Sub-branch description	%	Branch
E – Distribute unsold produce for free	43%	Act once products are in store → Act on storage → Empty stock
I – Sell products cheaper	13%	Act once products are in store → Through sales strategies → Target products
C – Reduce the quantity of products purchased by supermarkets	10%	Act before products arrive in store
U – Recycle/transform waste into non-edible product	9%	Act once products are in store → Transform product
S – Transform waste food into another food product	4%	Act once products are in store → Transform product

Note: Percentages are rounded to the nearest integer. For a full presentation of the C-K schema for this task, including all sub-branches, see Annex C1.
Source: Table B5.3

Figure 5.3 shows the relative share of ideas within these 5 sub-branches in select country-language groups. While the relative shares vary within each country language group, three things are notable. First, sub-branch E ("distribute unsold produce for free") represents the most common sub-branch in all country-language groups by a significant margin. Second, beyond sub-branch E, the distribution of ideas across the remaining sub-branches is relatively stable across country-language groups, with the cross-cultural variation remaining within 5 percentage points of the country-language average percentage in most cases (Table B5.3). This relative stability remains even when limiting the analysis to appropriate ideas only (i.e. those that achieve at least a score of 1 point for appropriateness), with only small changes to the overall distribution of ideas across all sub-branches (Table B5.4). Third, although the relative ordering of the sub-branches varies in each country-language group included in the study, particularly those that are second to fifth most common, it is notable that the five most common within-country idea nodes always correspond to one of these 5 sub-branches (Table B5.3 and Table B5.6). What these data show is that despite significant socio-cultural and linguistic diversity amongst students in the PISA CT Rescoring project sample, there is nonetheless broad homogeneity in the types of ideas that students suggest for this task.

Figure 5.3. Cross-cultural variability is relatively stable in *Food Waste*

Percentage of ideas within the top 5 most common sub-branches (on average across country-language groups), by select country-language groups



Note: The country-language groups with the largest and smallest shares of ideas within each of the 5 sub-branches are shown in the Figure, though not all country-language groups with available data are shown. Country-language groups are shown in alphabetical order following the average. The top 5 sub-branches are ordered by share of ideas at the average country-language level (highest to lowest). The max-min difference shows the percentage difference in the share of ideas between the country-language group with the highest share of ideas and the country-language group with the lowest share of ideas within each sub-branch. Percentages are rounded to the nearest integer.

Source: Table B5.3

Table 5.2 also describes the primary solution paths (or broader branch of the C-K Tree) that lead to the 5 most common sub-branches, on average across country-language groups. Notably, four of the five most common sub-branches sit within the same major branch of the C-K Tree ("act once fresh products are in the store"); this branch represents around 83% of all student ideas, on average across country-language groups, though it also encompasses several primary solution paths and the vast majority of the defined sub-branches within the solution space, including introducing sales strategies, transforming products into other edible or non-edible products, or acting on the storage of products by emptying stock or improving storage conditions (Figure 5.2). In contrast, around 10% of ideas corresponding to sub-branch C ("reduce the quantity of products purchased by supermarkets") – the third most frequently suggested sub-branch, on average – suggest addressing the problem from an alternative angle and solution path, that is taking action before the fresh products arrive at supermarkets. Very few students on average across country-language groups consider addressing the problem through other primary solution paths such as introducing some sort of regulatory framework or policies (around 3% ideas) or by reimagining the constraints of the problem itself (less than 1%).

Overall, these distributions of ideas across country-language groups for the *Food Waste* task imply two things that, at first, may seem contradictory. First, the high concentration of ideas within a select few solution paths, sub-branches and even specific nodes suggests that many students, across many country-language groups, think similarly when asked to come up with different ideas for addressing the problem of food waste in supermarkets. Second, the data suggest that beyond these few obvious solution ideas, students' ideas become quite varied. Indeed, 10 of the 25 sub-branches in the full C-K Tree schema contained between 1-10% of all ideas each, with a further 3% of ideas remaining unattributed to an existing category (Table B5.3).

These two phenomena may co-exist given the characteristics of the *Food Waste* task: there is a relatively large problem and solution space for students to explore; and the task is broadly accessible and familiar in some way to most students, meaning both that students will have a shared knowledge about broad solution approaches already implemented by supermarkets across contexts (e.g. giving away food at the end of the day or reducing prices), but that students may also have been directly exposed to various different or culturally-specific approaches in their own socio-cultural context or lived experience. For example, while sub-branch E ("distributing unsold produce for free") was consistently and significantly the most common sub-branch of solution ideas in all country-language groups, with "giving food directly to others for free" the most common instantiation (i.e. node) of this group of ideas in most country-language groups (around 23% of all ideas, on average), in Saudi Arabia, it was not even in the top 5 most common nodes; instead, a sizeable proportion of students (11%) in Saudi Arabia referred to creating a lottery system where people could win food (Table B5.6). Thus, while the broad solution approach remains similar, differences in the specific implementation of ideas might reflect differences in socio-cultural norms, values and practices.

5.1.2. The Exhibit: a harder scientific problem-solving task

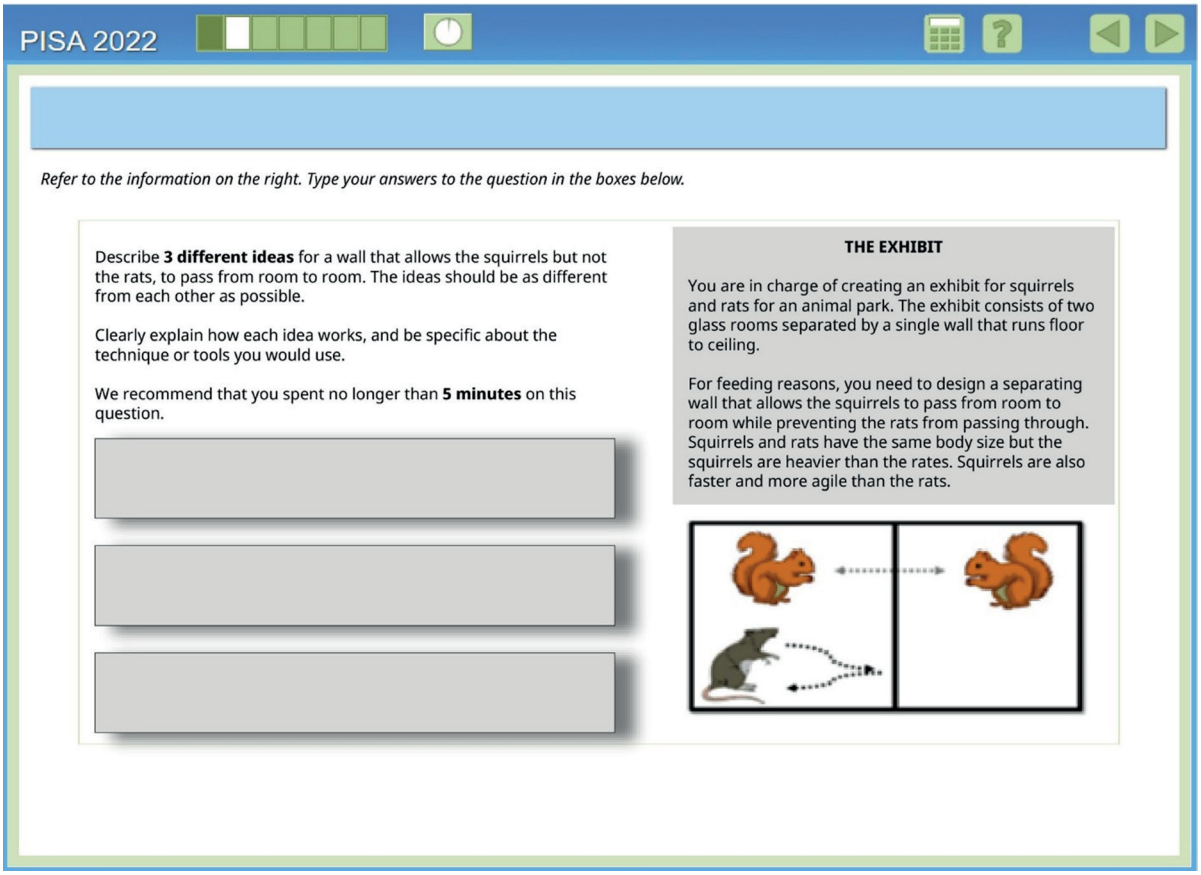
Do we observe a similar diversity and spread of ideas in a different problem-solving task included in the PISA CT Rescoring study? The C-K Tree scoring method was also applied to Task 7 – *The Exhibit*. This task was classified as a scientific problem-solving task in the PISA 2022 Creative Thinking test as it combines elements of physics, biology and engineering design. The task asks students to propose design ideas for constructing a separating wall in an animal park exhibit that would allow squirrels to pass from one room to the other for feeding purposes, while preventing rats in the enclosure from doing so (Figure 5.4). The item specified that while the squirrels and the rats have the same body size, the squirrels are heavier, faster and more agile than the rats.

The Exhibit is a comparatively harder task than *Food Waste*. In *The Exhibit*, students must combine their understanding of the characteristics of the two animals, as described in the task scenario, with some basic understanding of physics and engineering principles. It thus is more demanding from a domain knowledge point of view, and, unlike the *Food Waste* task, students may have limited lived experience with or exposure to such problems. *The Exhibit* also generally has a higher cognitive load for students (more text to read and understand, and greater demands on the level of detail required in responses). Indeed, *The Exhibit* appeared the most difficult of the seven items included in the PISA CT Rescoring project in terms of student

success in generating creative ideas, with a mean idea sum score of just 2.8 score points (on average across country-language groups) compared to a mean idea sum score of 3.7 score points on average across all tasks (Table B1.17 and Table B2.2). The criteria-based scores for response flexibility also indicate that students found it relatively difficult to generate multiple different ideas for this task, with 45% of students achieving a score of 0 for response flexibility and fewer than 1 in 3 students able to provide two sufficiently different ideas (Table B5.7).

The relative difficulty of *The Exhibit* task compared to the *Food Waste* task can also be viewed through the comparative openness of the problem and solution spaces. Although the C-K Tree for *The Exhibit* has 7 primary solution paths (branches), compared to the 8 primary solution paths in *Food Waste,* each path is less expansive: the full C-K Tree schema for *The Exhibit* includes just 13 sub-branches and 57 pre-defined end nodes, roughly half of those that were included in the equivalent for *Food Waste* (see Annexes C1 to C4).

Figure 5.4. *The Exhibit* task



Source: PISA 2022 Creative Thinking test.

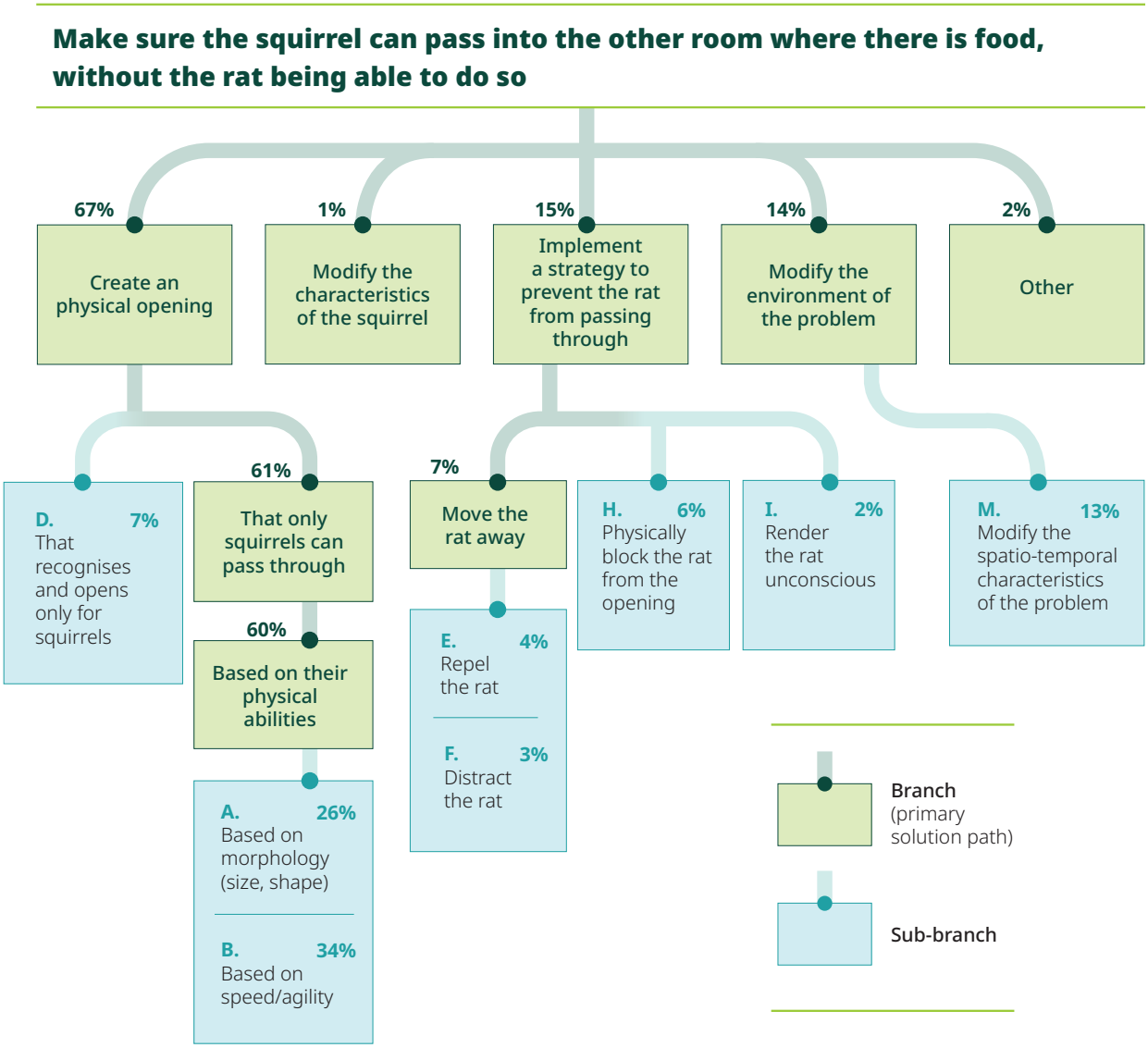
Figure 5.5 shows that the majority of student ideas across country-language groups are split between two sub-branches of the C-K Tree in *The Exhibit*: creating an opening that allows squirrels to pass through "due to their speed/agility" (34% all ideas; sub-branch B) or "due to their morphology" (25% all ideas; sub-branch A). Like in *Food Waste*, it is unsurprising that a significant share of all ideas corresponds to the same sub-branches given the task asks students to come up with different ideas of any quality. It could also be expected that the two most common ideas relate to characteristics of the squirrels that are directly referenced in the task prompt, given prior research on the effect of activating knowledge and providing examples on constraining flexibility in divergent thinking tasks (Agogué et al., 2014[9]; Smith, Ward and Schumacher, 1993[7]).

One difference in the distribution of ideas in *The Exhibit* task with respect to the *Food Waste* task is that ideas are relatively more evenly distributed across the different sub-branches and nodes that populate the solution

space. Most ideas in the *Food Waste* task are captured by just a few individual nodes, with a large spread of ideas across the remaining sub-branches and nodes in the solution space. In contrast in *The Exhibit,* student ideas were less concentrated within a specific node – both across and within sub-branches. For example, Table 5.3 shows that although the 5 most common nodes in *The Exhibit* task account for a similar proportion of all ideas as in the *Food Waste* task (between 61-62%), student ideas are more evenly distributed across these 5 nodes in *The Exhibit.* Table 5.4 also shows how the modal nodes by country-language group vary within the two most popular sub-branches (on average).

Figure 5.5. Over half of all ideas in *The Exhibit* correspond to two sub-branches of the C-K Tree

Percentage of student ideas within each branch and select sub-branches, on average across country-language groups



Note: Only sub-branches with over 2% of all student ideas on average across country-language groups are shown. Percentages shown in the figure are rounded to the nearest integer. The full C-K Tree schema for this task is detailed in Annex C3.

Source: Table B5.8

Table 5.3. Three out of the five most common ideas in *The Exhibit* focus on designs related to the speed or agility of the squirrels

Percentage of all student ideas within the 5 most common nodes (on average across country-language groups) on the C-K Tree for *The Exhibit*

Node description	%	Sub-branch (node)
Create an elevated path or course for the squirrel to traverse to the other side	16%	B (bf)
Separate the rats and squirrels permanently (i.e. avoid the problem altogether)	13%	M (mc)
Create an elevated opening that only the squirrel can jump to	13%	B (be)
Create an opening that opens with a balance/pressure plate on which the squirrel positions itself	11%	A (ae)
The opening is accessible only with the speed of the squirrel	8%	B (bb)

Note: Percentages are rounded to the nearest integer. For a full description of all nodes for this task, see Annex C4.
Source: Table B5.10

Table 5.4. The most common nodes within sub-branches A and B vary significantly by country-language group

Modal nodes by country-language group, within the two sub-branches with the largest share of student ideas (on average across country-language groups)

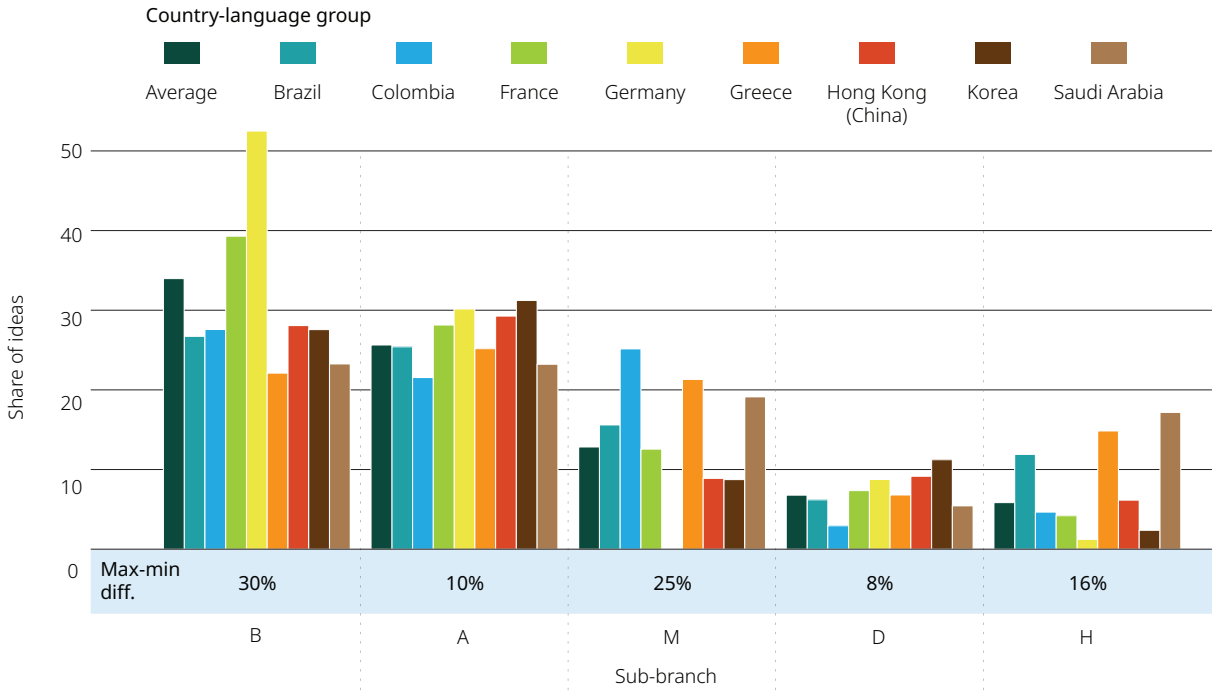
Sub-branch A: An opening for the squirrel based on its morphology			Sub-branch B: An opening for the squirrel based on its speed/agility		
Modal node	Node description	Country-language groups	Modal node	Node description	Country-language groups
ab	Opening is a small hole that only squirrels can pass through	Saudi Arabia	bb	Opening only accessible by speed	Hong Kong (China, *Chinese speaking*), Korea
ac	Make a tunnel opening	Colombia	be	Create an elevated opening (squirrels must jump to access)	Brazil, Canada (English provinces), France, Greece, Saudi Arabia
ad	An object (e.g. door) that only squirrels have the strength to push past	Hong Kong (China, *English speaking*)			
ae	A balance/pressure plate that recognises squirrels	Belgium (French community), Brazil, Canada (English and French provinces), Chile, France, Greece, Hong Kong (China, *Chinese speaking*), Korea, Slovenia	bf	Create an elevated path that squirrels must traverse to the other side	Belgium (French community), Canada (French provinces), Chile, Colombia, Germany, Hong Kong (China, *English speaking*), Slovenia
ag	Counterweight system	Germany			

Note: For a full description of all nodes for this task, see Annex C4.
Source: Table B5.11

In general, the cross-cultural variability in the distribution of ideas across sub-branches is greater in *The Exhibit* task compared to the *Food Waste* task. Figure 5.6 shows the share of ideas in select country-language groups within the 5 sub-branches that contain the largest shares of ideas (on average across country-language groups). The percentage of ideas varies significantly within sub-branch B ("creating an opening that allows squirrels to pass through due to their speed/agility"), ranging from 22% of all ideas in Greece to 52% ideas in Germany (Table B5.8). Although sub-branch B remains the most common sub-branch in 13 of the 16 country-language groups, the most common idea node in around half of the country-language groups corresponds to a different sub-branch (sub-branch A "an opening that only squirrels can pass through due to their size" or M "modify the spatial-temporal characteristics of the problem") (Table B5.11). In general, there is significant variation across country-language groups in the relative shares of ideas in four of the five most common sub-branches (A, B, M and H), with the relative shares of ideas in several country-language groups exceeding 5 percentage points of the country-language average in several instances (Table B5.8). The distribution of ideas across sub-branches also changes significantly when limiting the analysis to appropriate ideas only (i.e. those that achieve at least a score of 1 point for appropriateness), with a much greater concentration of ideas within sub-branches A and B on average across country-language groups (Table B5.9).

Figure 5.6. There is significant cross-cultural variability in the share of ideas across sub-branches in *The Exhibit*

Percentage of ideas within the top 5 most common sub-branches (on average across country-language groups), by select country-language groups



Note: The country-language groups with the largest and smallest shares of ideas within each of the 5 sub-branches are shown in the Figure, though not all country-language groups with available data are shown. Country-language groups are shown in alphabetical order following the average. The top 5 sub-branches are ordered by share of ideas at the average country-language level (highest to lowest). The max-min difference shows the percentage difference in the share of ideas between the country-language group with the highest share of ideas and the country-language group with the lowest share of ideas within each sub-branch. Percentages are rounded to the nearest integer.

Source: Table B5.8

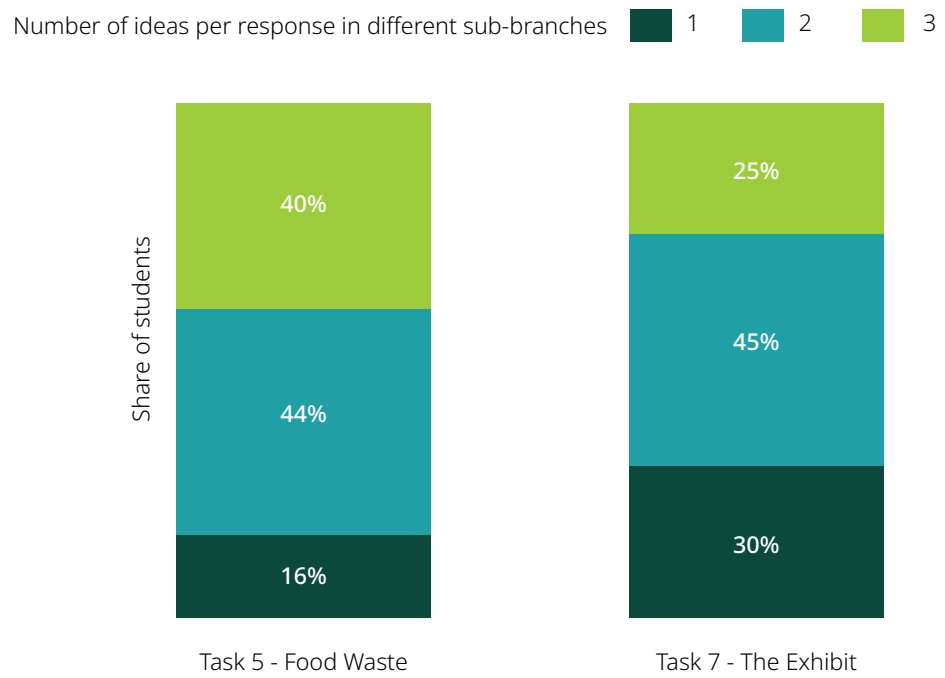
The comparative difficulty of *The Exhibit* task may influence these differences in the variation and distribution of ideas, likely because proposing functional solutions requires some degree of relevant knowledge from a range of disciplines. These differences in domain readiness across country-language groups may be reflected in the larger variation in modal nodes in *The Exhibit* task as well as the overall greater cross-cultural variation in the distribution of students' ideas. However, when considering only appropriate ideas suggested by students, the distribution of ideas is much more concentrated within the two sub-branches associated with the characteristics of the animals mentioned explicitly in the task prompt.

A sizeable proportion of students in *The Exhibit* task proposed ideas that significantly changed the problem presented (around 13%), for example by simply separating the two animals definitively and thus getting rid of the problem altogether (Table B5.8). In contrast, only 0.2% students proposed a similar problem-avoiding solution in the *Food Waste* task (Table B5.3). It may be that the harder difficulty of the problem, or the relative lack of on-topic solution ideas, caused students to think about ways to reframe the problem. While these types of "solutions" may well be original in the sense of adopting alternative perspectives, their appropriateness is questionable. However, in certain real-life situations, such lateral thinking and alternative perspective-taking may be useful.

The data from the C-K Trees also confirms that students found it relatively harder to think of different ideas in *The Exhibit* task. Figure 5.7 shows the percentages of students that suggested ideas in different sub-branches of the C-K Trees for both tasks. Although similar proportions of students were able to think of ideas corresponding to two different sub-branches of the C-K Tree in each task, many more students in *The Exhibit* task suggested multiple ideas within the same sub-branch compared to the same in the *Food Waste* task (30% students compared to 16%).

Figure 5.7. Students found it harder to suggest ideas in different sub-branches in *The Exhibit* than in *Food Waste*

Share of students that proposed ideas in one, two or three different sub-branches of the C-K Tree for the *Food Waste* and *The Exhibit* tasks (on average across country-language groups)



Note: Percentages are rounded to the nearest integer.
Source: Table B5.12 and Table B5.13

5.2. Are there differences in idea fixations across different student groups and socio-linguistic contexts?

For both the *Food Waste* and *The Exhibit* tasks, it is possible to define "fixation paths" and "expansion paths" (Box 5.2). Fixation paths refer to solutions mapped onto primary solution paths that many students follow, whereas expansion paths refer to solutions proposed outside of these most-followed solution paths. For the analyses in the rest of this chapter, we examine whether there are differences in the degree to which students generate ideas within fixation or expansion paths across different student groups and different socio-linguistic contexts.

Box 5.2. Identifying international fixation paths in students' thinking

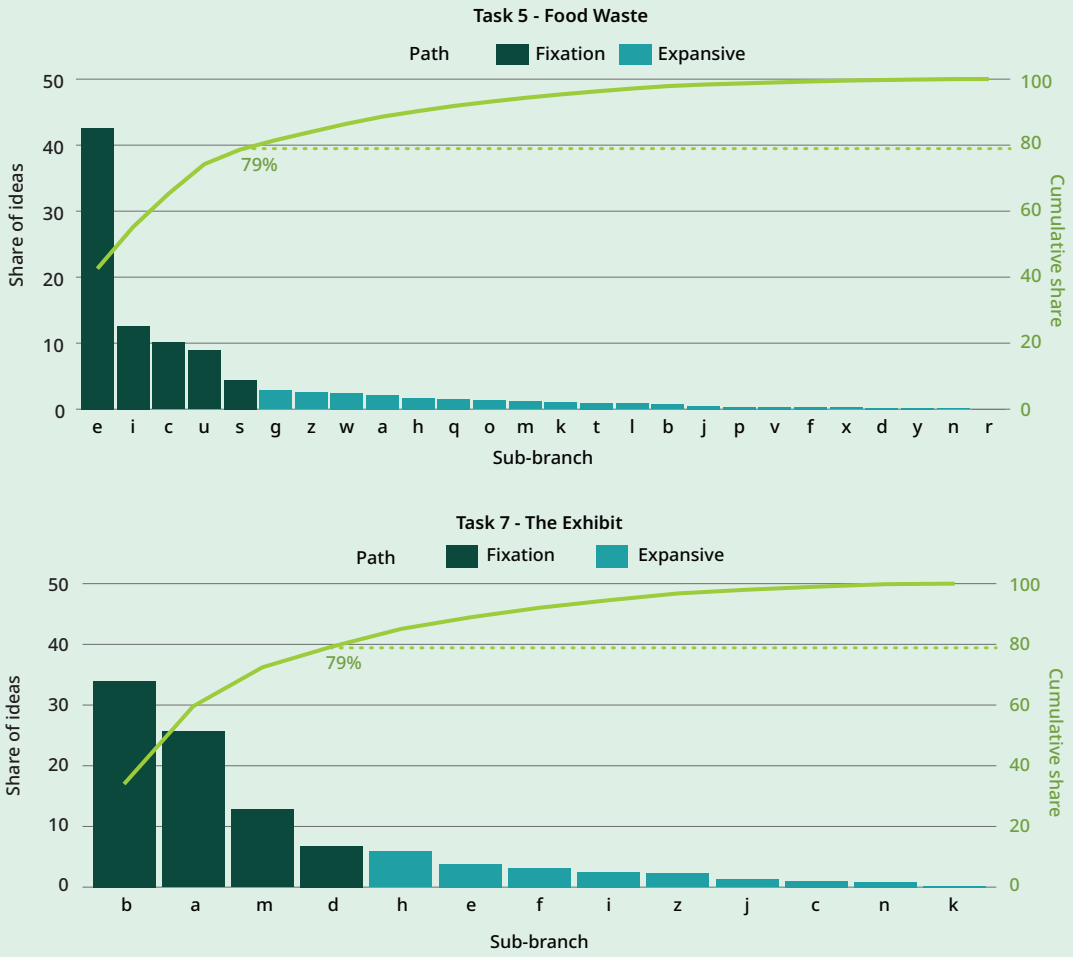
In previous research using the C-K method, "fixation paths" have been defined in the dataset as those that account for around 80% of all ideas in the sample (Agogué et al., 2014[9]).The fixation and expansion paths referred to in this chapter have been defined with respect to the international average, i.e. the sub-branches that account for the majority of all student ideas on each task, on average across country-language groups. Given the comparative distribution of the data across our two tasks, the sub-branches included in the fixation paths are those that cumulatively, from largest share to smallest share, include at least 79% of all student ideas on average across country-language groups. This represents the 5 most common sub-branches in *Food Waste* and the 4 most common sub-branches in *The Exhibit*.

Table 5.5. Sub-branches and cumulative frequency of ideas within the international fixation path for *Food Waste* and *The Exhibit*

Task 5 – Food Waste				Task 7 – The Exhibit			
#	Code	Sub-branch	Cum. Freq. (%)	#	Code	Sub-branch	Cum. Freq. (%)
1	E	Distribute unsold produce for free	43%	1	B	Create an opening that squirrels can pass through due to their speed/ agility	34%
2	I	Sell products cheaper	55%				
3	C	Reduce the quantity of products purchased by supermarkets	65%	2	A	Create an opening that squirrels can pass through due to their morphology	60%
4	U	Recycle/transform waste into a non-edible product	74%	3	M	Modify the spatial-temporal characteristics of the problem	72%
5	S	Transform waste into another food product	79%	4	D	Create an opening that recognises the squirrels	79%

Source: Table B5.14

Figure 5.8. Share of ideas in sub-branches within the international fixation path for *Food Waste* and *The Exhibit*



Note: Sub-branches are shown in descending order of the share of ideas at the average country-language group level.

Source: Table B5.3 and Table B5.8.

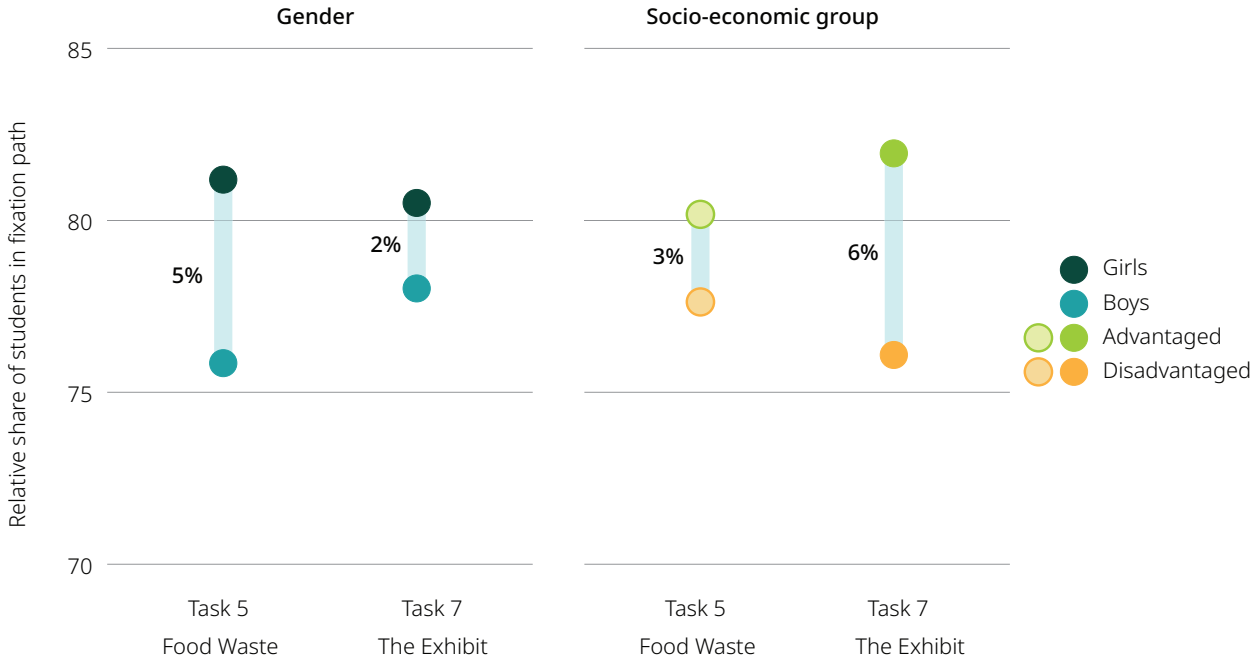
5.2.1. By gender and socio-economic status

Our data show that there are significant differences in idea fixation across student groups. Figure 5.9 shows that, in both tasks, girls are significantly relatively more likely than boys to propose solution ideas within the fixation path, on average across country-language groups. Advantaged students are also relatively more likely than disadvantaged students to suggest solution ideas within the fixation path, on average across country-language groups, though differences between advantaged and disadvantaged students are not significant in the *Food Waste* task. These patterns regarding relative differences in the ideation paths by gender and socio-economic status remain consistent even when replicating the analyses using nationally referenced fixation paths for each country-language group rather than the international fixation path, though differences between student groups tend to be smaller (Tables B5.15-B5.18).

These findings might seem counter-intuitive at first glance: how is it that girls and advantaged students – who performed better on the PISA 2022 creative thinking assessment (OECD, 2024[12]) and who tended to outscore boys and disadvantaged students in the PISA CT Rescoring project (see Chapters 3, 6 and 7) – also propose more ideas corresponding to the fixation path? What these results suggest is that creative ideas are not necessarily those that are "outside of the box" – at least, not in the types of everyday problem-solving tasks included in the PISA assessment. The PISA assessment tasks aimed to elicit "little c" expressions of creativity, rather than the types of "Big C" creative advances that might be more associated with out-of-the-box thinking. Ivcecic and Kaufman (2024[13]) underline the importance of task relevance for achieving "little-c" creative outcomes. In this way, it may be that "inside the box" ideas that capture some elements of originality are those that are most likely to be considered creative, rather than very remote or original ideas that only partially address the task.

Figure 5.9. Girls and advantaged students are more likely to suggest ideas within the fixation path than boys and disadvantaged students

Relative percentage of ideas in the within-task fixation path (on average across country-language groups), by gender and socio-economic status



Note: Significant differences between student groups are shown by filled markers. The relative percentage of ideas refers to the share of ideas amongst all student ideas within a given student group (i.e. amongst all girls, the percentage of ideas that correspond to sub-branches within the fixation path). The differences shown between student groups refer to the differences in relative percentages of student ideas within the fixation path and are rounded to the nearest integer. The fixation path is defined with respect to the international average (see Box 5.2).

Source: Tables B5.15-B5.18.

In both tasks, the sub-branches that contained the highest scoring ideas (on average) in country-language groups were not necessarily those that might be considered the most original types of ideas (based on frequency) amongst all the possibilities defined within the C-K Tree schema; however, they all clearly respond to the task goals in relatively practical ways (Table B5.19 and Table B5.20). For example, in *The Exhibit* task, sub-branches D ("an opening that recognises squirrels"), A ("an opening that only squirrels can pass through due to their morphology") and B ("an opening that only squirrels can pass through due to their speed/agility") are the modal sub-branches containing the highest mean idea sum scores (i.e. as either the highest-scoring

or second-highest scoring sub-branch) within each country-language group. In the *Food Waste* task, the equivalent modal sub-branches containing the highest mean idea sum scores within each country-language group are sub-branches C ("reduce the quantity of products purchased by supermarkets"), U ("recycle or transform waste into non-edible product") or S ("transform the produce into another food product"). In both tasks, these modal, high-scoring sub-branches all fall within the respective solution fixation paths.

Girls and advantaged students may be more likely to search for creative ideas "inside the box" for two reasons. Firstly, these student groups may be more willing or disposed to pay attention to and comply with the task goals, compared to boys and disadvantaged students, respectively. The analyses elsewhere in this report show that girls are significantly more likely to suggest appropriate ideas than boys, in general (see Chapter 6), and the PISA 2022 results found that boys tended to show more disengaged task behaviours than girls (OECD, 2024[12]). Not only do the fixation path sub-branches tend to correspond to appropriate classes of ideas to begin with, but girls also propose significantly more appropriate ideas than boys even within the fixation path in *Food Waste,* on average across country-language groups (Table B5.21). Similarly, advantaged students also propose significantly more appropriate ideas than disadvantaged students within the fixation path for both tasks, on average across country-language groups (Table B5.23 and Table B5.24).

Secondly, being able to provide more effective and relevant solutions inevitably requires domain-relevant knowledge. Both girls and advantaged students tend to outperform boys and disadvantaged students respectively in PISA, in general, suggesting that these latter groups may have more difficulty in understanding the task requirements or providing relevant, original solutions of quality in creative thinking tasks. Again, even amongst ideas within the fixation paths (so, those that are relatively common), advantaged students propose significantly more original and valuable ideas than their disadvantaged peers on average across country-language groups (Table B5.23 and Table B5.24). While there are no significant differences in the originality and value of ideas specifically within the fixation path between boys and girls, girls' overall idea sum scores are nonetheless significantly higher than boys' on average across country-language groups in both tasks (Table B5.21 and Table B5.22).

5.2.2. By socio-linguistic context

What about differences in idea fixations across socio-linguistic contexts? To what extent are our ideas and thinking patterns shaped by our cultural experiences and the language that we speak? Figure 5.10 and Figure 5.11 show the percentage of students' ideas that fall within the international fixation path for the two tasks, respectively, by country-language group, as well as by language group and geographic region (where these include at least two country-language groups that participated in the PISA CT Rescoring project).

As expected, the degree to which student ideas fall within the fixation path varies across different socio-linguistic groups. However, it's also notable that in most country-language groups, the fixation path does account for close to three quarters or more of all student ideas – again suggesting that student ideas across country-language groups for these two tasks are remarkably consistent. Of course, the fixation paths refer only to the sub-branches of solution pathways and not to specific nodes on the C-K trees, where there is larger variation in students' ideas, but it is striking that most students in most country-language groups follow the same general solution pathways. Part of this effect might be due to the nature of the PISA 2022 divergent thinking tasks – asking for students to generate different ideas of any quality, rather than focusing on coming up with different *creative ideas* – but the analyses in Chapter 2 of this report also demonstrated that students generally produced more creative ideas as a byproduct in these divergent thinking task types.

In both tasks, there are some relative outlier country-language groups. In Saudi Arabia, only around 67% of all student ideas fell within the international fixation path in the *Food Waste* task. Students were significantly more likely to suggest ideas related to the production of fresh food (i.e. grow less food in general or modify the type of food grown) and to the storage of fresh goods in supermarkets than on average across country-language groups in this task, and they were also significantly less likely to suggest ideas related to (re)distributing fresh food for free or transforming fresh food waste into other (non-consumable) products. Relatively fewer ideas for *The Exhibit* task were also captured by the fixation path for students in Saudi Arabia compared to the country-language average (difference of 8 percentage points).

Figure 5.10. At least two thirds of ideas within each country-language group corresponds to the international fixation path for the *Food Waste* task, though the share varies significantly

Percentage of student ideas within the international fixation path for *Food Waste*, by country-language group, language group and geographic region group

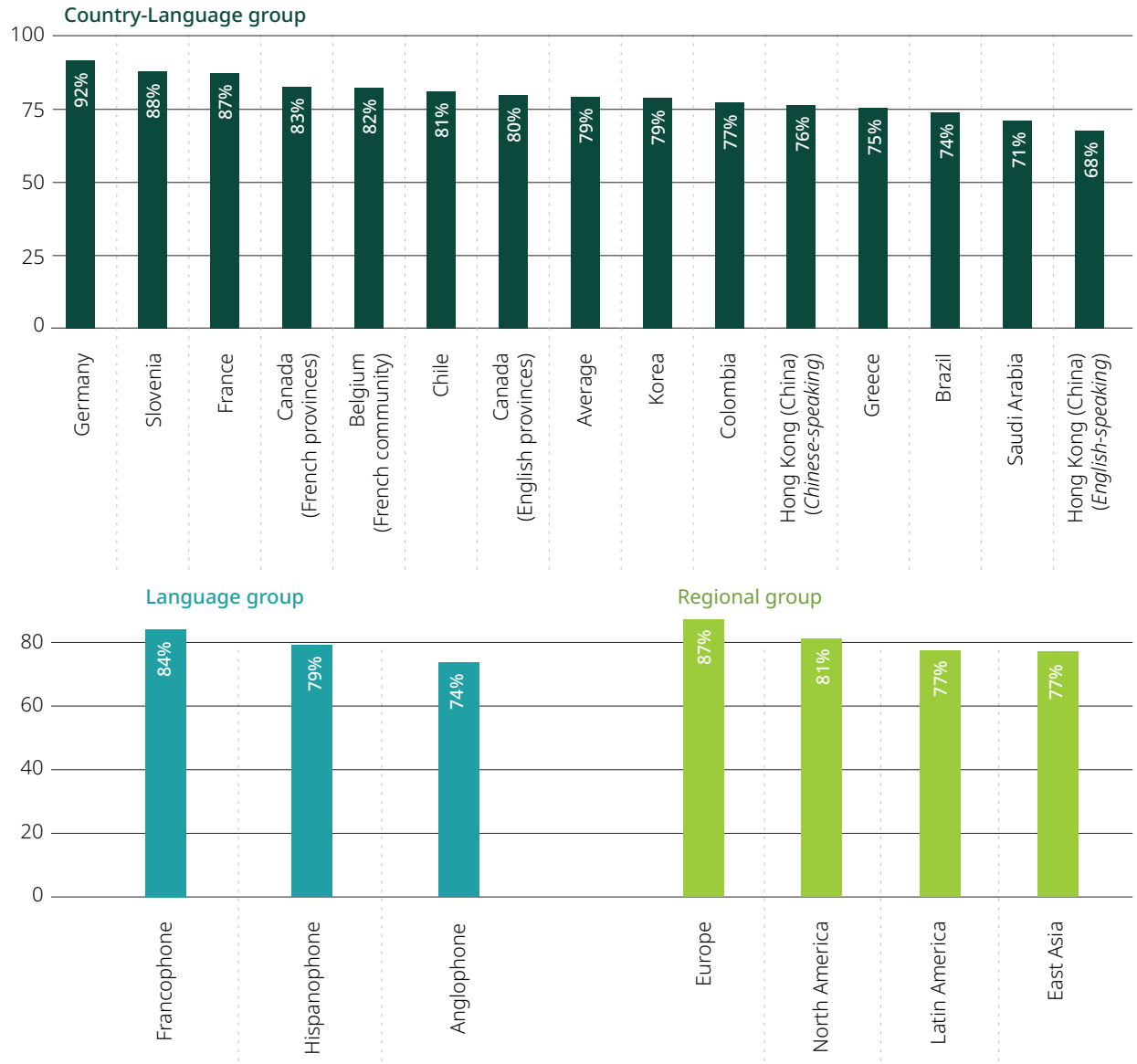
Note: Percentages are rounded to the nearest integer. Geographic region averages include at least two country-language groups. The sample size for Hong Kong (China, *English speaking*) is very small, so results for this country-language group should be interpreted with caution (for more information, see Annex A1). The country-language average is computed using Hong Kong (China, all), not Hong Kong (China, *English speaking*) and Hong Kong (China, *Chinese speaking*) as separate entities.

Source: Table B5.25

In Hong Kong (China, *English speaking*), only between 68-75% of all student ideas are within the fixation path for both tasks. These results may be an artefact of the small student sample size for the English-speaking student population in the sample; however, the Chinese-speaking sample of students from Hong Kong (China) also suggest fewer ideas within the fixation path than in most other countries, in both tasks. These results imply that students in Hong Kong (China) are particularly good at thinking flexibly in problem-solving tasks, or that there may be strong cultural differences in the primary solution paths that students in Hong Kong (China) follow compared to other country-language groups.

Figure 5.11. At least two thirds of ideas within each country-language group corresponds to the international fixation path for *The Exhibit* task, though the share varies significantly

Percentage of student ideas within the international fixation path for *The Exhibit*, by country-language group, language group and geographic region group



Note: Percentages are rounded to the nearest integer. Geographic region averages include at least two country-language groups. The sample size for Hong Kong (China, *English speaking*) is very small, so results for this country-language group should be interpreted with caution (for more information, see Annex A1). The country-language average is computed using Hong Kong (China, all), not Hong Kong (China, *English speaking*) and Hong Kong (China, *Chinese speaking*) as separate entities. There are no data for Portugal for this task.

Source: Table B5.26

Key findings and implications

- In general, students do not think much outside of the box when asked to come up with different ideas to everyday problem-solving tasks – both within and across country-language groups
  - In both problem-solving tasks, many students in many countries follow relatively fixed solution paths. In the *Food Waste* task, over 4 in 10 ideas corresponded to the same sub-branch of the C-K Tree, on average, and in both tasks, the 5 most common nodes accounted for over 60% of all student ideas on average across country-language groups.
- There was greater cross-cultural variability in students' ideas in *The Exhibit* task compared to the *Food Waste* task, with ideas more evenly spread across nodes (in general) and greater variability in the relative shares of ideas within different sub-branches.
  - Part of this greater cross-cultural variability may be due to the task difficulty and differences in students' prior knowledge and experience across relevant disciplines (e.g. biology, physics, engineering).
- Knowledge matters in supporting creative problem-solving: it is important to teach students how to apply creative thinking in domain-relevant ways, enabling them to find appropriate and practical solutions with an original twist
  - Students were far more likely to propose ideas that ignored the task requirements or that avoided the problem altogether in *The Exhibit*, again perhaps due to the greater difficulty of the task and a relative lack of familiar solution approaches from which to suggest appropriate ideas.
- It may neither be reasonable nor desirable to expect students to think outside of the box in these types of everyday problem-solving contexts: successful "everyday" creativity might be more about thinking inside the box in creative ways (Ivcevic and Kaufman, 2024[13]).
  - Girls and advantaged students seem to be more successful at thinking creatively "inside the box", most likely due to differences across student groups in task engagement, domain-relevant knowledge, and being able to identify more appropriate solutions.

References

Adamson, R. (1952), "Functional fixedness as related to problem solving: A repetition of three experiments.", Journal of Experimental Psychology, Vol. 44, pp. 288-291. [5]

Agogué, M. (2015), "Resisting classical solutions: The creative mind of industrial designers and engineers", Psychology of Aesthetics, Creativity, and the Arts, Vol. 9/3, pp. 313-318, [8]

Agogué, M. et al. (2014), "The impact of type of examples on originality: explaining fixation and stimulation effects", Journal of Creative Behavior, Vol. 48/1, pp. 1-12, https://doi.org/10.1002/jocb.37. [9]

Cassotti, M. et al. (2016), "Fixation effect in creative ideas generation: Opposite impacts of example in children and adults", Thinking Skills and Creativity, Vol. 19, pp. 146-152, https://doi.org/10.1016/j.tsc.2015.10.008. [10]

Defeyter, M. and T. German (2003), "Acquiring an understanding of design: evidence from children's insight problem solving", Cognition, Vol. 89/2, pp. 133-155, https://doi.org/10.1016/s0010-0277(03)00098-2. [6]

Hatchuel, A. and B. Weil (2009), "CK design theory: An advanced formulation.", Research in Engineering Design, Vol. 19, pp. 181-192. [11]

Ivcevic, Z. and J. Kaufman (2024), "Inside the box: Considering the role of appropriateness for creativity", Psychology of Aesthetics, Creativity and the Arts, https://doi.org/10.1037/aca0000703. [13]

Jansson, D. and S. Smith (1991), "Design fixation", *Design Studies*, Vol. 12/1, pp. 3-11, https://doi.org/10.1016/0142-694X(91)90003-F. [3]

OECD (2024), *PISA 2022 Results (Volume III): Creative Minds, Creative Schools*, PISA, OECD Publishing, https://doi.org/10.1787/765ee8c2-en. [12]

Purcell, A. and J. Gero (1996), "Design and other types of fixation", *Design Studies*, Vol. 17/4, pp. 363-383, https://doi.org/10.1016/S0142-694X(96)00023-3. [2]

Smith, S., T. Ward and R. Finke (eds.) (1995), Fixation, incubation and insight in memory, problem-solving, and creativity, MIT Press. [1]

Smith, S., T. Ward and J. Schumacher (1993), "Constraining effects of examples in a creative generation task", *Memory and Cognition*, Vol. 21, pp. 837-845. [7]

Viswanathan, V. and J. Linsey (2011), *Understanding fixation: A study on the role of expertise*. [4]

06

Gender differences in creative thinking: are girls really better than boys at thinking creatively?

One of the most striking and consistent findings described in the PISA 2022 report was that boys did not perform better than girls, on average, in any of the 64 countries and economies that participated in the creative thinking assessment (OECD, 2024[1]). In fact, girls scored nearly 3 points higher than boys on average in the PISA creative thinking scale across OECD countries (a large performance difference). Even after accounting for students' performance in reading, girls performed significantly better than boys with similar reading scores in around half of all countries and economies. These performance differences in favour of girls persisted across all types of task in the PISA test, including those targeting different ideation processes and applications in different domain contexts.

Gender differences in creative thinking may be rooted in broader cognitive, psychological and behavioural patterns. Prior research on creativity has found that girls tend to outperform boys in specific creative tasks, especially those involving associative thinking and elaboration (Baer and Kaufman, 2008[2]). While boys tend to perform better in divergent thinking tasks (Kazemian et al., 2024[3]; Baer and Kaufman, 2005[4]; Kim, 2006[5]), girls tend to score higher on originality and verbal creativity assessments. Socio-cultural norms and practices may contribute to these findings by influencing the development of boys' and girls' creative self-concept and reinforcing gendered cognitive styles and problem-solving strategies (Kim, 2006[5]; Bem, 1981[6]). Neuroimaging studies have also found gender differences in cognition: men and women activate different brain regions when solving creative tasks — particularly those requiring divergent thinking — suggesting the existence of gender-specific neural pathways (Abraham et al., 2013[7]). Other studies suggest that male advantage in creativity is influenced by personality traits such as openness to experience, assertiveness, and risk-taking (Feist, 1998[8]; Byrnes, Miller and Schafer, 1999[9]), although the PISA 2022 data found that 15-year-old girls generally reported higher levels of creative self-efficacy (especially in expressive or artistic tasks), and significantly stronger attitudes typically associated with supporting creative work including openness to experience and imagination and adventurousness (OECD, 2024[1]).

Overall, prior research on gender differences in creativity has found mixed results, with various factors including cognitive, motivational, affective and socio-cultural norms at play. One reason for these mixed findings in the research may be due to developmental differences between boys and girls across these factors, with periods of change and stability occurring at different ages (Slobodskaya, 2021[10]; Lau and Cheung, 2010[11]; He, 2018[12]). While the PISA 2022 results offered some hypotheses for why such strong gender differences in performance were observed in the PISA creative thinking test – namely differences in test engagement between boys and girls, and girls' propensity to hold more positive attitudes and beliefs related to creativity – another contributing factor may have been the PISA scoring method itself. The PISA 2022 creative thinking scoring approach prioritised identifying sufficiently different or original responses rather than evaluating the holistic creative quality of students' ideas. Will more granular scoring methods, such as those applied in the PISA CT Rescoring project (see Box 6.1), offer different insights into gender differences in creative thinking performance?

Box 6.1. Which score data are used in this chapter?

The analyses in this chapter draw upon two types of rescored data:

- Standardised holistic creativity scores: Each student response was scored based on judges' evaluations of the overall creative quality of the response. The holistic creativity score reflects a relative ranking of creative quality compared to all other responses in that country-language-task group. To account for differences in the distribution of scores across country-language-task groups, the raw holistic creativity scores (range 1-7 score points) were standardised and transformed to a scale that ranges from 1-10 score points, with a mean of 5 score points.
- Criteria scores: Each idea within a student response was scored for its appropriateness, originality and value. A cumulative idea sum score was also produced by aggregating the scores across these three criteria. Tasks classified as 'generate diverse ideas' tasks required students to submit multiple ideas per response; for these tasks, each student response was also scored for flexibility (i.e. the number of different ideas in the response).

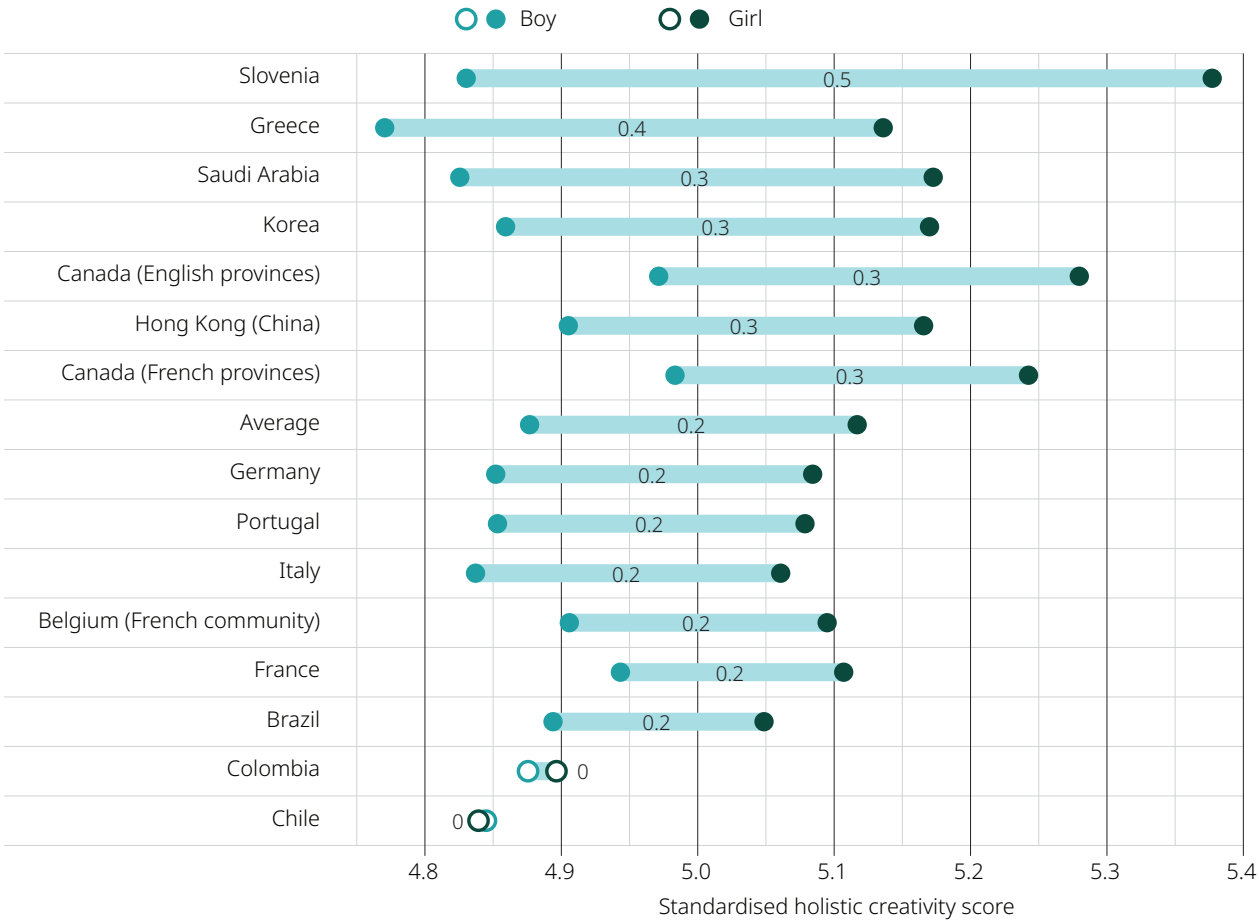
The analyses in this chapter focus on gender differences in scores resulting from these two methods. For more information on the scoring methods, see Annex A2.

6.1. Do gender differences in creative thinking persist across different scoring methods and across country-language contexts?

Across all country-language groups included in the PISA CT Rescoring study, girls achieved a higher mean standardised holistic creativity score than boys. Comparing the scores of boys and girls, girls scored 5.1 points on average whereas boys scored 4.9 points – a modest but statistically significant difference (0.2 points) in performance (Figure 6.1). The difference in mean score between boys and girls was also statistically significant in 13 of the 15 country-language groups included in the study, with the largest gaps observed in Slovenia (0.6 points), Greece (0.4 points), Saudi Arabia, Korea, Canada (both English and French provinces) and Hong Kong (China) (all 0.3 points). In Slovenia, this equates to over one third of a standard deviation in creative thinking performance between boys and girls – see Box 6.2. Only in Chile and Colombia was there no statistically significant difference in the performance of girls and boys in terms of their standardised holistic creativity scores. The PISA 2022 results also found no statistical differences in the performance of boys and girls in creative thinking in Chile, and a significant but below average difference in creative thinking performance in favour of girls in Colombia (OECD, 2024[1]).

Figure 6.1. In no country-language group do boys significantly outscore girls on average

Mean standardised holistic creativity scores of boys and girls (all tasks), by country-language group



Note: Significant differences between boys and girls are shown by filled markers, and values indicating the mean score difference have been rounded to the nearest decimal point. Standardised holistic creativity scores have a mean of 5 score points for each country-language-task group. Country-language groups are ranked in descending order of the difference in mean standardised holistic creativity scores between boys and girls. Mean standardised holistic creativity scores are computed within each country-language group on the pooled responses across all tasks.

Source: Table B6.1

Box 6.2. How big of a score difference?

The standardised holistic creativity scores in each country-language-task group were centred around a mean of 5 score points, with a standard deviation of 1.5 score points. This means that a standardised holistic score difference of 0.5 score points between boys and girls equates to a third of a standard deviation in performance (Table B6.2).

As for the criteria-based scores used in this chapter, the standard deviation in students' appropriateness scores was 0.8 score points on average across country-language groups. For the originality and value criteria, the standard deviation in students' scores were 0.9 score points, respectively, on average across country-language groups. For all three criteria, a score difference of 0.3 score points equates to around a third of standard deviation in performance, on average – although scores in some country-language groups varied more than others (Table B6.3).

Comparing gender differences in the PISA CT Rescoring project with gender differences in PISA 2022

In general, gender differences in performance observed in the PISA CT Rescoring project are similar to or smaller than those observed in the PISA 2022 results for both creative thinking and reading (OECD, 2023[13]; OECD, 2024[1]). Table 6.1 summarises the gender effect size in each test context.

Table 6.1. Gender effect sizes in creative thinking and reading

Gender differences expressed as a proportion of the standard deviation in performance (all students), by country-language group*

	PISA CT Rescoring project	PISA 2022 Creative Thinking	PISA 2022 Reading
Belgium (French community)	0.1	0.2*	0.3*
Brazil	0.1	0.2	0.2
Canada (English provinces)	0.2	0.2*	0.2*
Canada (French provinces)	0.2		
Chile	0.0	0.1	0.1
Colombia	0.0	0.1	0.1
Germany	0.2	0.3	0.2
France	0.1	0.2	0.2
Greece	0.2	0.3	0.3
Hong Kong (China)	0.2	0.4	0.2
Italy	0.1	0.2	0.2
Korea	0.2	0.3	0.3
Saudi Arabia	0.2	0.5	0.4
Portugal	0.2	0.2	0.2
Slovenia	0.4	0.4	0.5

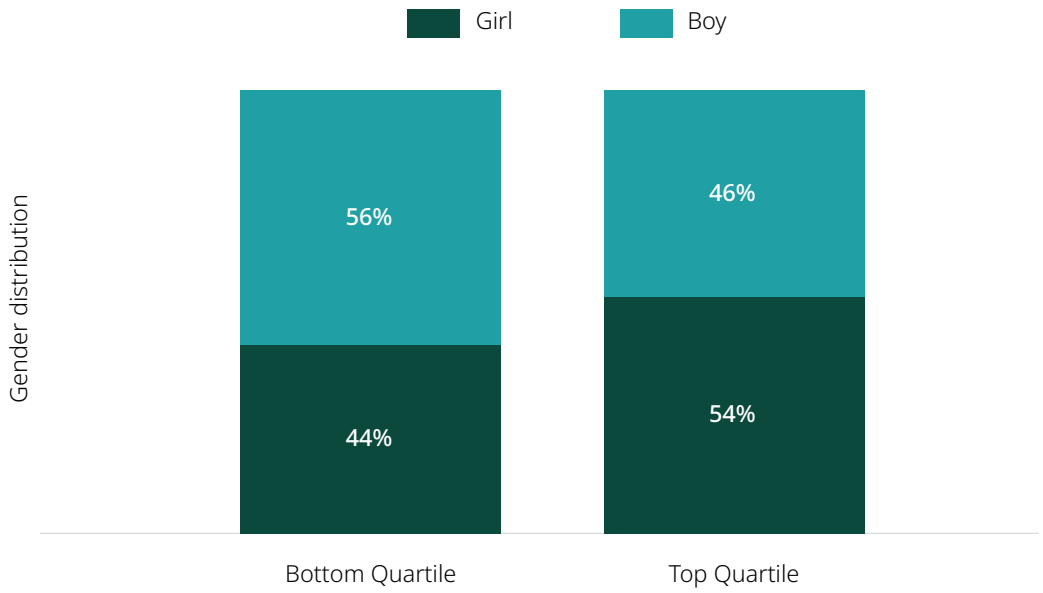
Note: *Gender effect sizes for the PISA 2022 creative thinking assessment and reading assessment are computed at the country level for Canada and Belgium, not at the country-language group level.
Effect sizes are rounded to the nearest decimal point.

Source: Table B6.4

Looking at the top 25% and bottom 25% of performers across country-language-task groups also reveals significant gender differences. Figure 6.2 shows that, on average, 54% of students within the top quartile of performers are girls compared to 46% boys. In Brazil, over 20 percentage point more girls than boys are top performers in the tasks included in the study, and in Greece, Canada (French provinces), France, Portugal and Belgium (French community), the gender difference is over 10 percentage points in favour of girls (Table B6.5). However, in nine country-language groups, there are no significant differences between boys and girls within the top 25% of performers across all tasks, suggesting that the performance of boys and girls is more variable towards the lower end of the score scale. Indeed in the bottom quartile, the gender differences are much starker: close to 56% of students in the bottom 25% of performers are boys, on average across country-language groups, compared to just 44% of girls. In Slovenia, nearly 65% of the lowest quartile of performers are boys (a gender difference of around 30 percentage points), and in Saudi Arabia, the figure is around 63% (a difference of 26 percentage points).

Figure 6.2. There are relatively more girls than boys amongst top performers, and relatively many more boys than girls amongst low performers

Percentage of boys and girls within the top/bottom 25% of standardised holistic creativity scores per task (pooled responses across tasks), on average across country-language groups



Note: Differences in the share of boys and girls amongst responses within the bottom/top quartile per task (pooled across tasks) are statistically significant. The standardised holistic creativity scores were jittered, i.e. a random small digit (1e-6) was added to each score, so that equal quartiles in each country-language-task group could be produced. Values in the figure have been rounded to the nearest integer.

Source: Table B6.5

A second scoring method applied in the PISA CT Rescoring study involved scoring different dimensions of students' ideas separately (see Box 6.2). These dimensions included idea appropriateness, originality and value. Additionally, judges scored each response to 'generate diverse ideas' tasks for idea flexibility (i.e. how different were the multiple ideas provided by students). Our results show that girls scored higher than boys, on average, in all three idea score criteria (appropriateness, originality and value) to a small but significant degree, representing a difference of around one-tenth of the mean standard deviation in each of the score criteria (Table B6.6). While most country-language groups observed significant gender differences in at least one of the idea score criteria, there was substantial variation across country-language groups (see Box 6.3). Gender differences in the mean idea sum scores of boys and girls were more consistent across country-language groups, with significant results observed in all but four country-language groups (Brazil, Canada [English provinces], Chile and Colombia) and an average performance difference of 0.3 score points (or

around one-third of the mean standard deviation in performance) in favour of girls. As for response flexibility, girls also showed a small but significant score advantage compared to boys, on average. Most notably, when considering the gender differences observed across the different criteria-based scores collectively, boys did not score significantly higher than girls in any of the idea or response score criteria, on average, in any of the country-language groups in the study.

Box 6.3. Cross-cultural variation in gender differences across the different dimensions of creativity

Criteria-based scores for each idea within a response

Appropriateness

Girls scored significantly higher than boys in idea appropriateness in 10 of the 15 country-language groups. Although significant, these differences generally remain small (around 0.1 score points on average), with the largest differences observed in Slovenia, Saudi Arabia and Korea (difference of 0.2 score points). No significant differences in this score criteria were observed in Belgium (French community), Brazil, Chile, Colombia and France.

Originality

Fewer gender differences across country-language groups were observed in the mean idea originality of boys and girls compared to the other score criteria. While girls scored 0.1 points more than boys (a small but significant difference) on average across country-language groups, a similarly significant relationship was observed in only six country-language groups: Belgium (French community), Canada (French provinces), Saudi Arabia, Italy, Portugal and Slovenia. In Slovenia, the mean score point difference between boys and girls was close to 0.2 points in favour of girls.

Value

Around half of the country-language groups showed small and significant differences in mean idea value scores between boys and girls. These differences were significant and around 0.1 score points in Belgium (French community), France, Greece, Korea, Saudi Arabia and Portugal, and largest in Slovenia and Germany (around 0.2 score points).

Idea sum scores

Much larger score differences were observed between boys and girls with respect to their mean idea sum scores (i.e. their aggregated appropriateness, originality and value scores). These differences were significant in 11 of the 15 country-language groups, and on average across country-language groups, girls scored around 0.3 score points higher than boys overall. In some cases, differences in the mean idea sum scores of boys and girls are as large as 0.5 score points in Germany and Korea, 0.6 score points in Saudi Arabia, or even 0.7 score points in Slovenia.

Criteria-based scores for each response

Flexibility

Around half of the country-language groups also showed small, significant differences in mean response flexibility scores between boys and girls. Girls in Belgium (French community), France, Greece, Korea, Saudi Arabia and Portugal scored around 0.1 score points more than boys in idea flexibility, on average, and in Slovenia and Germany, they scored on average 0.2 score points more than boys.

Note: Mean idea and response scores are computed within each country-language on the pooled responses across all tasks.

Source: Table B6.6

Taken together, the analyses using data from the two scoring methods indicate the persistence of gender differences in creative thinking on the PISA test. It seems that no matter how creative performance is evaluated, girls consistently and significantly outperform boys on average across most country-language groups – even if differences in some of the scores are generally small in magnitude. Other factors might explain some of this apparent gap in performance, for example differences in test engagement between boys and girls (Box 6.4), but the PISA CT Rescoring study data confirms that the gender differences observed in the PISA 2022 results were not simply an artefact of the PISA 2022 scoring method.

Box 6.4. Gender differences in engagement with the PISA 2022 creative thinking test

In the context of the PISA 2022 creative thinking assessment, task engagement emerged as a critical factor that may contribute to explaining some of the large observed gender differences in performance. The PISA report showed that boys displayed higher levels of disengaged behaviours than girls across a variety of task engagement indicators — a trend that was consistent across most PISA participating countries and economies (OECD, 2024[1]). In particular, boys were far more likely to engage in "rapid responding" behaviours compared to girls, on average. Boys also showed higher levels of disengagement than girls with all types of task in the assessment, on average, although gender differences in task engagement were relatively smaller in the scientific problem-solving domain than in the other three task domains. These differences in task engagement would non-discriminately affect the scores of boys and girls in any scoring method applied to the PISA creative thinking data.

Task engagement is not necessarily construct-irrelevant in the context of creative thinking: confluence approaches of creativity underline that engagement is an important factor that enables individuals to produce creative work. However, whether these differences in task engagement in the PISA creative thinking test would translate to similar differences outside of a testing environment remains an open question. For example, one study found boys tend to exhibit more disengaged behaviours in low-stake test environments (Leng and von Davier, 2025[14]). The gender differences observed in the PISA 2022 results and in this report should therefore be interpreted within this context.

6.2. Are gender differences in creativity more pronounced in certain domains?

While gender differences in creative thinking skill generally appear consistent across the diverse national-linguistic contexts in the PISA CT Rescoring study, is this also the case when examining the performance of boys and girls across the different types of task included in the study? Prior research and reviews on gender differences in creativity measures have found mixed results (Baer and Kaufman, 2008[2]; Alabbasi et al., 2025[15]; Taylor et al., 2024[16]) with many studies using relatively general or abstract divergent thinking tasks and few studies focused on exploring gender differences using domain-specific creativity measures. Some such studies have found no significant differences in the performance of men and women in figural (visual) and creative writing tasks (Taylor and Barbot, 2021[17]), or small performance advantages for girls in storytelling and figural tasks (Ivcevic et al., 2022[18]).

The data from the PISA CT Rescoring study show that, on average across country-language groups, girls score significantly higher than boys in their standardised holistic creativity scores in the written expression, visual expression and social problem-solving tasks included in the study. In these three task domains, girls scored around 0.3 score points higher than boys (around one-fifth of a standard deviation), on average. However, the score gap is markedly smaller for tasks in the scientific problem-solving domain (difference of

0.1 score points on average) (Table B6.7). Overall, these domain-level gender differences in holistic creativity scores are generally consistent with the strengths and weakness of boys and girls reported in the PISA 2022 results (OECD, 2024[1]).

A more nuanced picture emerges when examining the mean scores of girls and boys by domain context within each country-language group. Figure 6.3 shows that girls tend to score significantly better than boys in the written expression and social problem-solving tasks in over half of the country-language groups participating in the study, respectively, compared to only four country-language groups in the single visual expression task included in the study. Interestingly, in the four country-language groups where girls do score significantly better than boys in the visual task (Canada [French provinces], Greece, Korea and Slovenia), the performance difference is large (between a third and a half of the mean standard deviation in performance) – and typically more so than most differences observed in the written or social problem-solving tasks across country-language groups. While the PISA 2022 results found a more consistent difference in the performance of boys and girls in the visual expression items on the PISA test, the fact that the PISA CT Rescoring project included only one visual item may explain this discrepancy.

For tasks in the scientific problem-solving domain, we observe a different general pattern in gender differences in student scores. In most country-language groups, and on average, there is no significant difference in the scores of boys and girls in these tasks: only in Saudi Arabia and Canada (English provinces) do girls still score significantly higher than boys in their standardised holistic creativity scores in scientific problem-solving. The opposite is true in Chile, where boys score significantly better than girls in scientific problem-solving – the only instance in our study where boys significantly outperform girls in a subset of tasks.

One reason for the different results in the scientific problem-solving tasks may be gendered differences in engagement with these items, relative to the other tasks in the PISA test. As described in Box 6.4, gender differences in task engagement persisted across all task types in the PISA creative thinking test, although they were relatively smaller in the scientific problem-solving tasks. It is not that boys are more engaged with scientific problem-solving tasks specifically – in fact, both boys and girls exhibited higher levels of relatively rapid responding and non-responding behaviours to the scientific problem-solving tasks compared to the other task types – but rather, girls show higher levels of relative disengagement than boys with the scientific problem-solving tasks compared to all other tasks (OECD, 2024[1]).

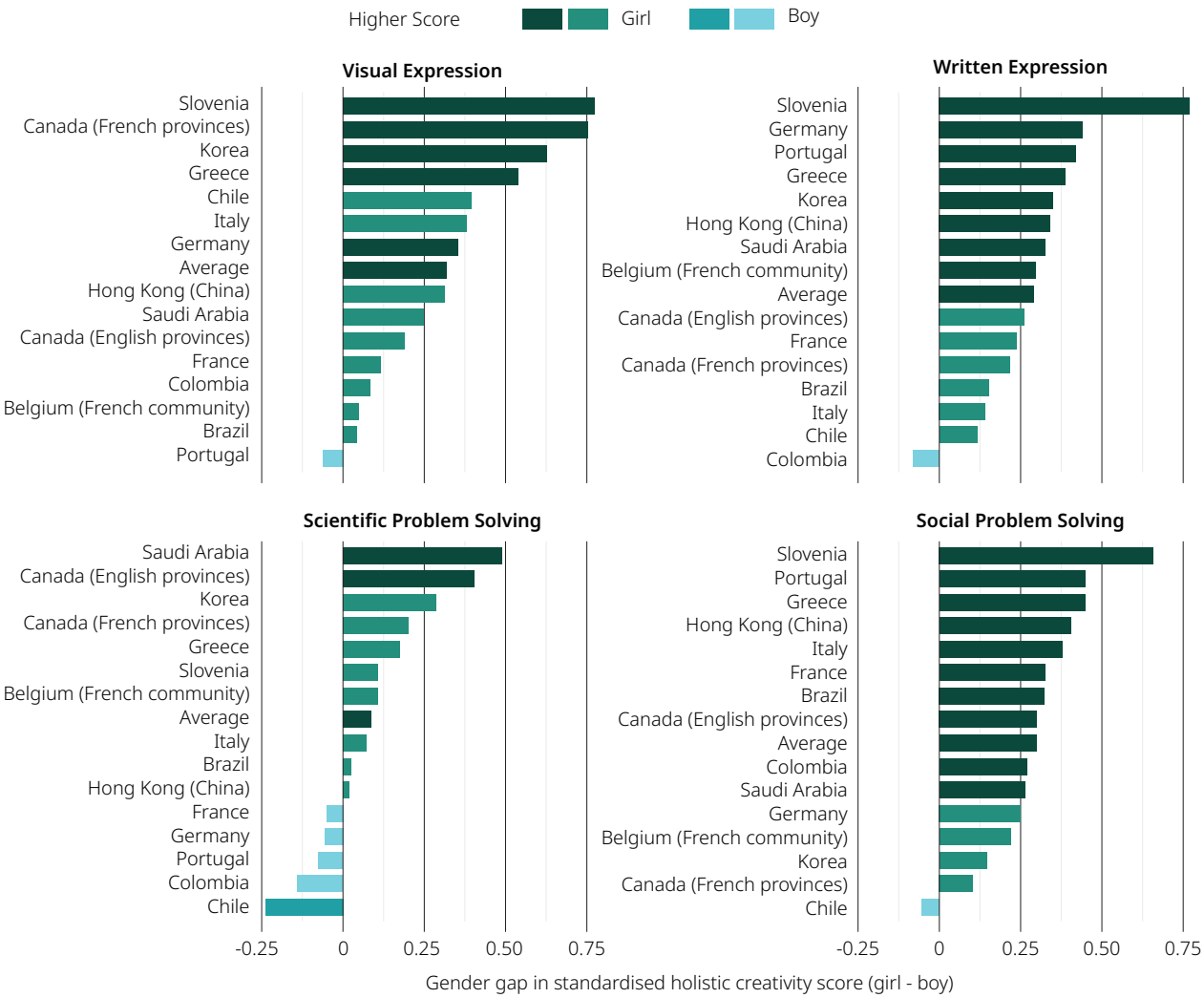
The relative disengagement by girls with creative scientific problem-solving tasks may stem from gender biases in the perception and attribution of creativity across different domains. Research has found that individuals typically consider some creative domains to be more suitable for women, such as crafts and literature, while other domains are more suitable for men, including science and engineering (Taylor and Barbot, 2023[19]). The same study also found that stereotypically "masculine-agentic" traits (e.g. adventurous, ambition, competitive, courageous, daring, decisive, self-reliant and willing to take risks) were rated as significantly more important for creativity in the science domain than stereotypically "feminine-communal" traits (e.g. cooperative, helpful to others, nurturing, sensitive, supportive, sympathetic, understanding of others, and warm in relations with others); in contrast, there were no significant differences in the perceived importance of the two groups of traits in the literature or crafts domains, despite gender biases in attributions of creativity in favour of women in these two domains. In other words, regardless of the gender of the individual, not only does creativity in scientific contexts tend to be viewed as a male domain, but stereotypically male traits are considered particularly beneficial for creativity in scientific contexts.

These more general gender biases related to creativity in science mirror previous PISA findings that have found that girls report lower science self-efficacy than boys (OECD, 2019[20]). Are these findings also replicated in students' creative self-efficacy beliefs, specifically girls' and boys' attitudes towards engaging in creative tasks in different domains? Figure 6.4 shows the mean difference in girls' and boys' self-reported creative self-efficacy in select tasks, by country-language group. On average, girls report feeling more confident than boys at being creative, in general, with large differences (around 10 percentage points or more) observed in Italy, Saudi Arabia and France. Girls also tend to score higher in the overall index of creative self-efficacy, on average across country-language groups (Table B6.8). When it comes to telling creative stories, the gender differences in favour of girls are particularly pronounced, with over 5 percentage point more girls than boys agreeing with the statement on average, and over 10 percentage point more girls than boys agreeing in Saudi Arabia, Slovenia, Greece and Germany. Yet when asked about their confidence in thinking creatively in science and invention contexts, girls are notably less confident – and even less confident than boys,

on average. For example, many more boys than girls feel confident that they can think of many good ideas for science experiments (difference of 10 percentage points on average), and more boys than girls also report feeling confident to invent new things (difference of around 6 percentage points on average). In around half of all country-language groups, gender differences in creative self-efficacy related to thinking of good science experiments are significant and relatively large in favour of boys, with the difference around or greater than 15 percentage points in Greece, Hong Kong (China), Belgium (French community) and France, and over 20 percentage points in Portugal and Korea. Despite overall greater creative performance and creative self-efficacy than boys, girls nonetheless seem far less likely than boys to feel confident about being creative in scientific or engineering-type contexts. These findings therefore align with previous research on gender biases in how individuals perceive creative work and creative individuals in different domains (Taylor and Barbot, 2023[19]).

Figure 6.3. Girls significantly outscore boys in tasks in all domains, on average, and in most country-language groups

Mean difference in standardised holistic creativity scores between boys and girls, by task domain and by country-language group

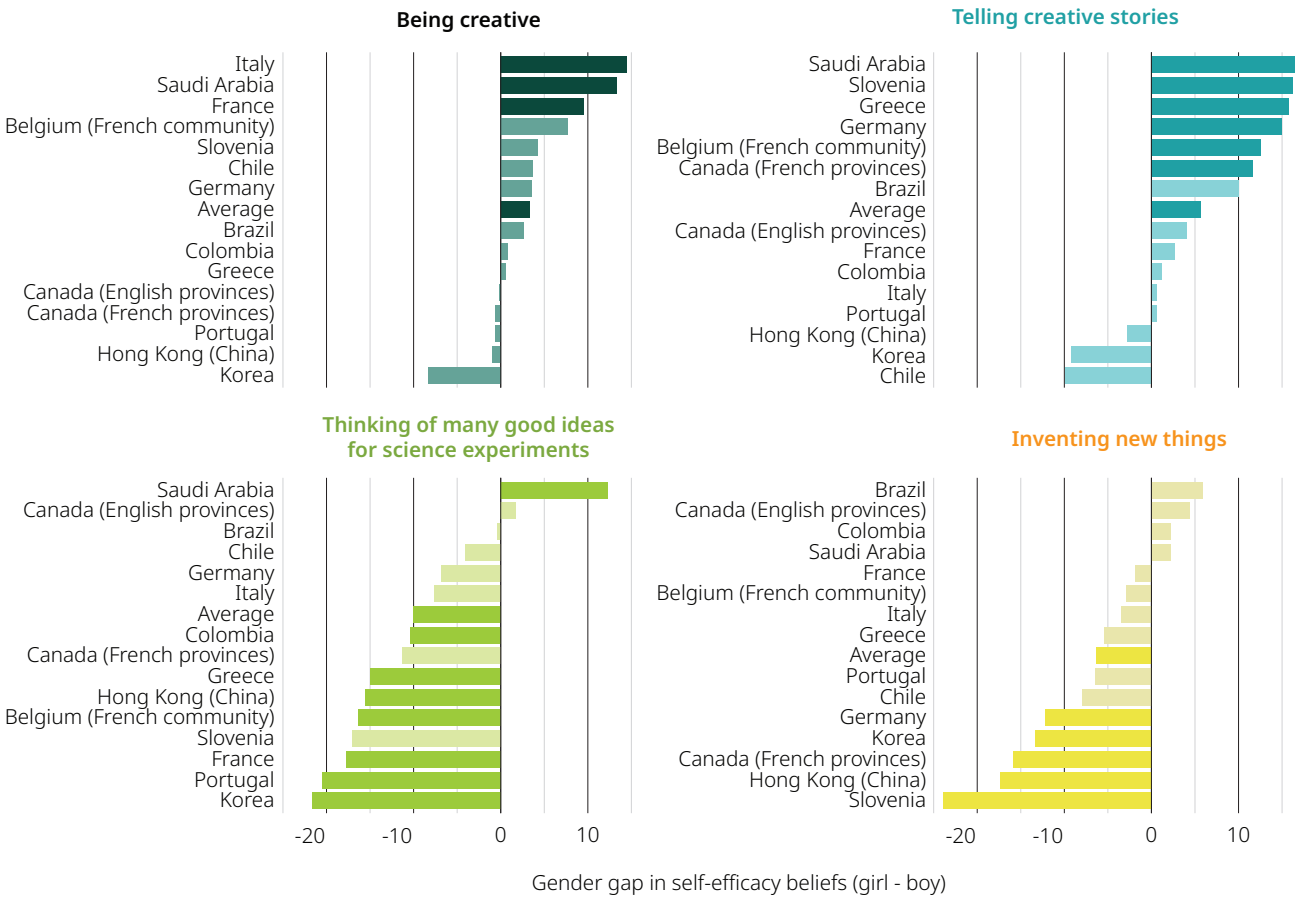


Note: Significant values are shown in colours of a darker shade for boys/girls, respectively. Country-language groups are ranked in descending order of the mean difference in standardised holistic creativity score (highest to lowest). Mean standardised holistic creativity scores of boys and girls are computed within each country-language on the pooled responses across relevant tasks in each domain.

Source: Table B6.7

Figure 6.4. Girls are more confident than boys at being creative, in general, but report far less confidence than boys in creative scientific or invention tasks

Mean difference in the percentage of boys and girls that reported feeling confident or very confident with select items from the creative self-efficacy index, by country-language group



Note: Significant differences are shown in colours of a darker shade for each statement. Country-language groups are ranked in descending order of the mean difference in percentage of boys and girls (girl - boy) that reported feeling confident or very confident with each statement (highest to lowest).

Source: Table B6.8

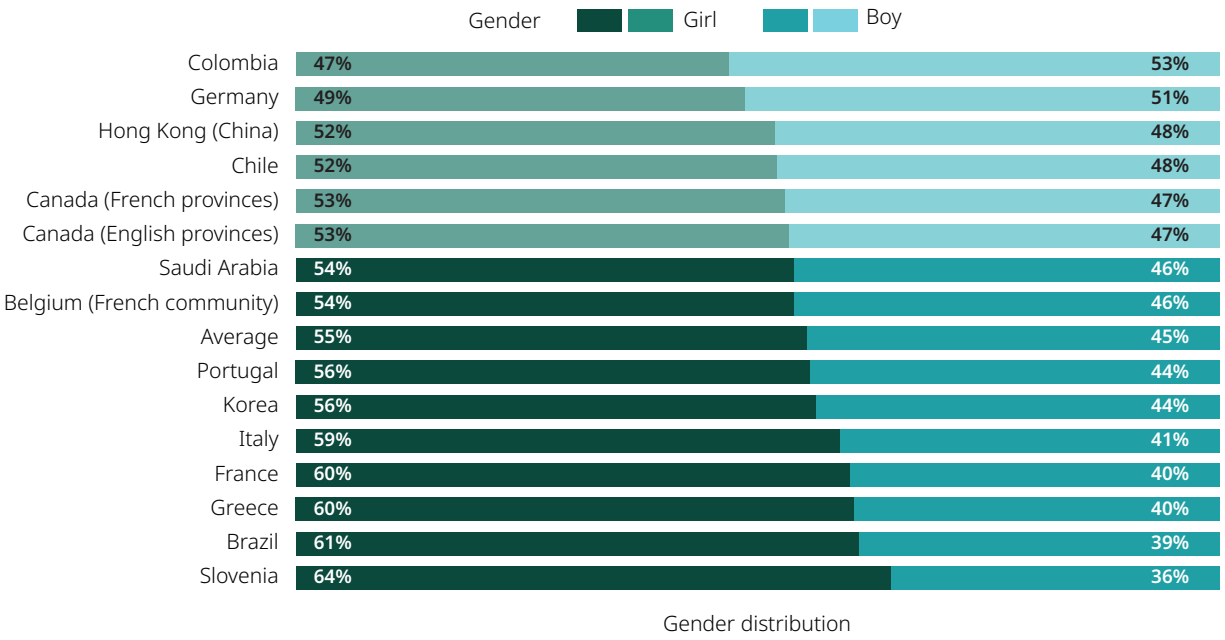
6.3. Do the data support the greater male variability hypothesis in creative thinking?

One hypothesis often explored in educational psychology is the Greater Male Variability Hypothesis (GMVH), which suggests that boys exhibit greater variability than girls in general across a range of performance contexts. Do our data confirm this hypothesis in the context of creative thinking? Results from our study show that, on average across country-language groups, there was marginally more variation in the standardised holistic creativity scores amongst boys than girls, with the largest differences in score variability observed in Saudi Arabia and Korea (Table B6.2). However, in most country-language groups, differences in the variation between boys and girls tended not to be statistically significant.

The GMVH suggests that boys are more likely to be represented at both high and low score extremes. Figure 6.5 shows the share of girls and boys amongst students within the 90th percentile of standardised holistic creativity scores in each task within each country-language group. In most country-language groups and on average, the share of girls amongst students within the 90th percentile, represented by the proportion of the bar in green, exceeds the share of boys amongst students within the 90th percentile, represented by the proportion of the bar in blue. In nine country-language groups, these gender differences are statistically significant. In Brazil, Greece and France, the gap in the share of girls and boys amongst the top 10% of performers in tasks is between 20-22 percentage points, and in Slovenia this difference reaches 29 percentage points.

Figure 6.5. On average, girls are overrepresented amongst the top 10% of performers in creative thinking tasks

Percentage of boys and girls amongst students in the 90th percentile of standardised holistic creativity scores per task (pooled responses across tasks), by country-language group



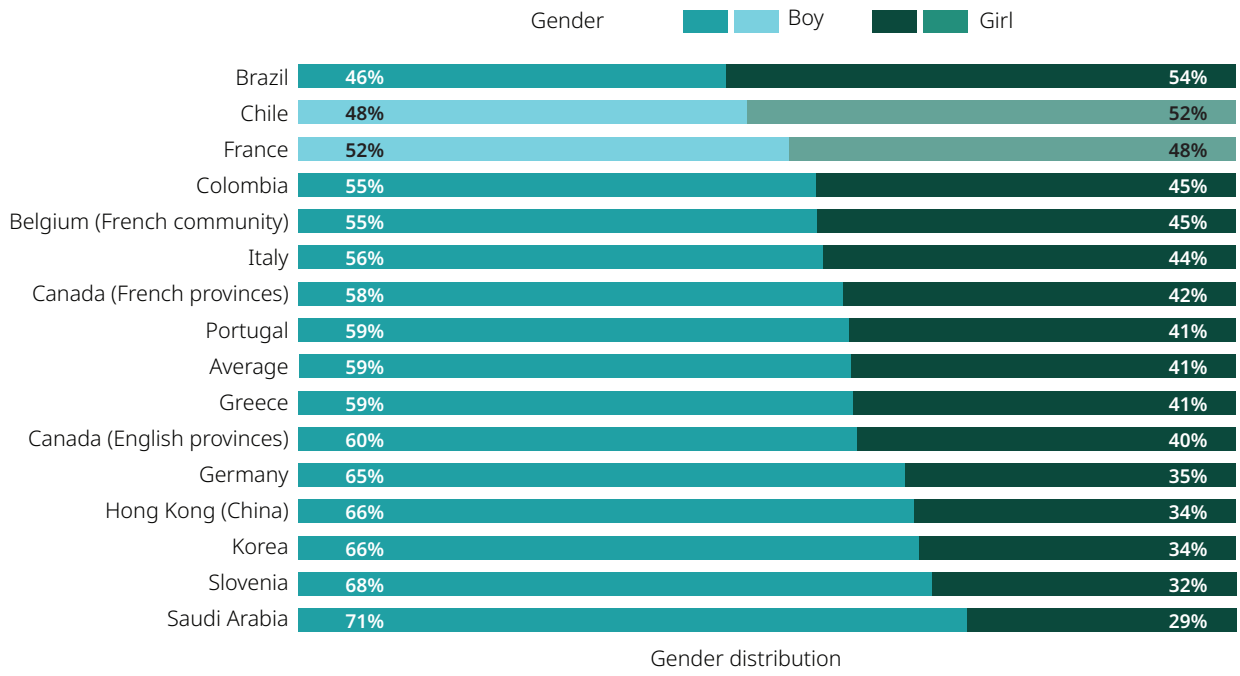
Note: Significant differences are shown in colours of a darker shade for boys/girls, and values are rounded to the nearest integer. Country-language groups are ranked in ascending order of the percentage of girls (lowest to highest). The standardised holistic creativity scores were jittered, i.e. a random small digit (1e-6) was added to each score, so that equal deciles in each country-language-task group could be produced. The country-language average is computed amongst all responses within the top decile per task (pooled across tasks).

Source: Table B6.9

When examining the same amongst the lowest 10% of performers on tasks (i.e. 10th percentile), boys are significantly and consistently overrepresented – with 59% of students amongst the lowest 10% of scorers across tasks being boys. Boys are particularly overrepresented in this group in Saudi Arabia (71% students), Slovenia (68% students), Korea (66% students), Hong Kong (China) (66% students) and Germany (65% students) (Table B6.9). The only country-language group where girls are statistically significantly overrepresented amongst the lowest performers in tasks is Brazil, where 54% are girls.

Figure 6.6. On average, boys are overrepresented amongst the lowest 10% of performers in creative thinking tasks

Percentage of boys and girls amongst students in the 10th percentile of standardised holistic creativity scores per task (pooled responses across tasks), by country-language group



Note: Significant differences are shown in colours of a darker shade for boys/girls, and values are rounded to the nearest integer. Country-language groups are ranked in ascending order of the percentage of boys (lowest to highest). The standardised holistic creativity scores were jittered, i.e. a random small digit (1e-6) was added to each score, so that equal deciles in each country-language-task group could be produced. The country-language average is computed amongst all responses within the bottom decile per task (pooled across tasks).

Source: Source: Table B6.9

The findings in this section provide limited support for the greater male variability hypothesis in creative thinking – though it could be argued that rather than boys exhibiting a significantly greater variation in performance than girls per se, the distribution of boys and girls differs mostly in that there is a higher incidence of low performers among boys than there are high performers. Again, these findings may be explained as much by gender differences in engagement with the PISA test (Box 6.4) as they are by the GMVH.

Alternatively, the overrepresentation of boys at the lower end of the holistic creativity scores may reflect broader gender differences in engagement with creative tasks, which have been argued to stem from certain notions of masculinity that run counter to the tacit attribution and promotion of many creative domains as a feminine pursuit (Taylor and Barbot, 2023[19]; Scholes and Nagel, 2010[21]). The OECD's project on fostering critical and creative thinking examined the role of emotions while learning, and offers practitioners several ideas for differentiated engagement strategies that might help to support boys' interest in and motivation to engage in creative thinking (OECD, 2022[22]).

Key findings and implications

- Girls consistently score higher than boys in the PISA creative thinking tasks.
  - This pattern was observed consistently across different country-language groups, task types and scoring methods, suggesting it is a relative stable finding that is not driven by any single scoring method, task feature, or sample anomaly.
  - However, the gender effect size observed in the PISA CT Rescoring data tended to be similar to or smaller than the equivalent gender effect size observed in both the PISA 2022 creative thinking and reading results. It may be that the more granular holistic scoring method used in the Rescoring study mitigates gender differences to some extent, for example by rewarding more subtle differences in idea quality and/or by being less driven by differences in test engagement.
- The only domain context in which girls did not show a significant and moderate performance advantage over boys in creative thinking is in scientific problem solving – although girls still significantly outperformed boys in the scientific problem-solving tasks, on average across country-language groups.
  - These results may rather reflect a relative weakness for girls in creative thinking, rather than a particular strength for boys. Indeed, data from the PISA 2022 assessment revealed that girls showed higher levels of relative disengagement than boys with the scientific problem-solving tasks compared to all other tasks.
  - Differences between boys and girls in creative thinking performance and task engagement in the scientific problem-solving tasks may reflect gendered attitudes and beliefs about creativity in science (Taylor and Barbot, 2023[19]). Girls reported greater levels of creative self-efficacy in general, but notably not in tasks in scientific or invention contexts.
  - Initiatives focused on supporting girls' participation and confidence in Science, Technology, Engineering and Mathematics (STEM) education may help to redress the relative weakness of girls in creative thinking in scientific contexts.
- Boys show slightly more variation in creative thinking performance than girls, but boys are more overrepresented amongst low performers than girls are overrepresented amongst high performers.
  - Efforts should focus on raising the performance of boys at the lower end of the creative thinking score distribution. Initiatives that help to encourage boys' engagement with creative tasks may be a productive strategy for practitioners (OECD, 2022[22]).

References

Abraham, A. et al. (2013), "Gender differences in creative thinking: behavioral and fMRI findings", *Brain Imaging and Behavior*, Vol. 8/1, pp. 39-51, https://doi.org/10.1007/s11682-013-9241-4. [7]

Alabbasi, A. et al. (2025), "Gender differences in creative potential: A meta-analysis of mean differences and variability", *Psychology of Aesthetics, Creativity and the Arts*, Vol. 19/1, pp. 87-100, https://doi.org/10.1037/aca0000506. [15]

Baer, J. and J. Kaufman (2008), "Gender differences in creativity", *The Journal of Creative Behavior*, Vol. 42/2, pp. 75-105, https://doi.org/10.1002/j.2162-6057.2008.tb01289.x. [2]

Baer, J. and J. Kaufman (2005), "Bridging generality and specificity: The amusement park theoretical (APT) model of creativity", *Roeper Review,* Vol. 27/3, pp. 158-163, https://doi.org/10.1080/02783190509554310. [4]

Bem, S. (1981), "Gender schema theory: A cognitive account of sex typing.", *Psychological Review*, Vol. 88/4, pp. 354-364, https://doi.org/10.1037/0033-295x.88.4.354. [6]

Byrnes, J., D. Miller and W. Schafer (1999), "Gender differences in risk taking: A meta-analysis", *Psychological Bulletin*, Vol. 125/3, pp. 367-383, https://doi.org/10.1037/0033-2909.125.3.367. [9]

Feist, G. (1998), "A meta-analysis of personality in scientific and artistic creativity", *Personality and Social Psychology Review*, Vol. 2/4, pp. 290-309, https://doi.org/10.1207/s15327957pspr0204_5. [8]

He, W. (2018), "A 4-year longitudinal study of the sex-creativity relationship in childhood, adolescence, and emerging adulthood: Findings of mean and variability analyses", *Frontiers in Psychology,* Vol. 9, https://doi.org/10.3389/fpsyg.2018.02331. [12]

Ivcevic, Z. et al. (2022), "Gender and creative ability: Mean differences and", *Thinking Skills and Creativity*, Vol. 46, pp. 1-12, https://doi.org/10.1016/j.tsc.2022.101186. [18]

Kazemian, N. et al. (2024), "Some effects of sex and culture on creativity, no effect of incubation", *Empirical Studies of the Arts*, Vol. 42/2, pp. 536-559, https://doi.org/10.1177/02762374231217638. [3]

Kim, K. (2006), "Can we trust creativity tests? A review of the Torrance Tests of Creative Thinking (TTCT)", *Creativity Research Journal,* Vol. 18/1, pp. 3-14, https://doi.org/10.1207/s15326934crj1801_2. [5]

Lau, S. and P. Cheung (2010), "Developmental trends of creativity: What twists of turn do boys and girls take at different grades?", *Creativity Research Journal*, Vol. 22/3, pp. 329-336, https://doi.org/10.1080/10400419.2010.503543. [11]

Leng, D. and M. von Davier (2025), "Gender differences and types of test-taking behaviors in PIRLS 2021", *Applied Measurement in Education*, Vol. 38/1, pp. 75-93, https://doi.org/10.1080/08957347.2025.2486032. [14]

OECD (2024), *PISA 2022 Results (Volume III): Creative Minds, Creative Schools,* PISA, OECD Publishing, Paris, https://doi.org/10.1787/765ee8c2-en. [1]

OECD (2023), *PISA 2022 Results (Volume I): The State of Learning and Equity in Education,* OECD Publishing, Paris, https://doi.org/10.1787/53f23881-en. [13]

OECD (2022), *How do girls and boys feel when developing creativity and critical thinking? Lessons for gender equity,* OECD Education Spotlights, OECD Publishing. [22]

OECD (2019), *Why don't more girls choose to pursue a career in science?,* PISA in Focus, 93, OECD Publishing. [20]

Scholes, L. and M. Nagel (2010), "Engaging the creative arts to meet the needs of twenty-first-century boys", *International Journal of Inclusive Education*, Vol. 16/10, pp. 969-984, https://doi.org/10.1080/13603116.2010.538863. [21]

Slobodskaya, H. (2021), "Personality development from early childhood through adolescence", *Personality and Individual Differences,* Vol. 172, pp. 1-7, https://doi.org/10.1016/j.paid.2020.110596. [10]

Taylor, C. and B. Barbot (2023), "Gender biases in attributions of creativity across domains", *Creativity Research Journal*, Vol. 37/3, pp. 395-405, https://doi.org/10.1080/10400419.2023.2285546. [19]

Taylor, C. and B. Barbot (2021), "Gender differences in creativity: Examining the greater male variability hypothesis in different domains and tasks", *Personality and Individual Differences,* Vol. 174, pp. 1-9, https://doi.org/10.1016/j.paid.2021.110661. [17]

Taylor, C. et al. (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,* Vol. 126/6, pp. 1161-1179, https://doi.org/10.1037/pspp0000484. [16]

07

Does socio-economic disadvantage hinder students' creativity?

The PISA 2022 results showed that, while everyone has the potential to be creative, differences in student socio-economic status may prevent many students from fully reaching their creative potential. On average, advantaged students scored close to one standard deviation higher than disadvantaged students in the PISA 2022 creative thinking test – enough to significantly decrease the overall mean performance of some countries and economies (OECD, 2024[1]). While student reading and mathematics scores explained a large proportion of the variation in students' creative thinking scores, there were nonetheless significant differences in the performance of advantaged and disadvantaged students in creative thinking even after accounting for student performance in the core domains (OECD, 2024[1]). In general, research into the association between socio-economic status and creativity has tended to find a small but significant positive relationship (Acar et al., 2023[2]; Castillo-Vergara et al., 2018[3]).

There are many reasons why socio-economic disadvantage may hinder students' creative potential. Beyond the well-documented economic, cultural and educational experiences and mechanisms that are known to affect student achievement overall, advantaged students may also have better access to diversified life experiences and activities that can nurture creative idea generation (Xu and Pang, 2020[4]; Jankowska, Lebuda and Gralewski, 2024[5]). Access to more diverse life experiences might support creative thinking in three ways: (1) by enabling students to develop a more heterogenous knowledge base from which to come up with creative ideas (Gocłowska and Crisp, 2014[6]); (2) by exposing students to new norms more frequently, in turn increasing cognitive flexibility (Ritter et al., 2012[7]); and (3) by increasing individuals' openness to experience and intellect, personality traits that have consistently demonstrated a strong association with creative performance (Batey and Furnham, 2006[8]; Zhang et al., 2018[9]).

On the other hand, socio-economic hardships might expose students more frequently to the types of everyday problem-solving situations that require creative thinking (Acar et al., 2023[2]). Economic disadvantage and/or membership of a minority culture could be considered developmental adversities that provide an alternative type of diversifying experience, in line with the "hidden talents" theory (Ellis et al., 2023[10]), and research has found that children facing difficult socio-economic and cultural circumstances can develop enhanced cognitive abilities as an adaptive response (Young et al., 2022[11]).

Overall, while economic privilege can facilitate access to certain diverse life experiences, not all are inherently costly: for example, bilingualism and multiculturalism has been found to be positively associated with creativity (Lee and Kim, 2011[12]). Other factors outside of access to new experiences might also contribute to the poorer performance of disadvantaged students in creative tasks, including lower levels of self-efficacy (Lu, Ding and Nie, 2024[13]; Karwowski, 2011[14]) or challenges with executive functions like cognitive flexibility, working memory and inhibition that are known to support cognitive performance across tasks (Kupczyszyn, Filippetti and Oros, 2024[15]).

What new insights can data from the PISA CT Rescoring project shed on the relationship between creative thinking performance and socio-economic status? Do socio-economic differences persist across different scoring methods, and which aspects of coming up with creative ideas do disadvantaged students struggle with most? Do these patterns persist across country-language groups and different types of task? And what role (if any) do factors like multilingualism, self-efficacy or immigrant background play in mediating socio-economic differences in creative thinking performance?

7.1. Which scoring criteria do disadvantaged students struggle with most?

As expected, results from the PISA CT Rescoring project find that student socio-economic background is significantly associated with their scores derived from both the holistic judgement method and the criteria-based method (Box 7.1). On average across country-language groups, students' standardised holistic creativity scores and mean response idea sum scores were positively correlated (r=0.2 and r=0.17, respectively) with their PISA economic, social and cultural (ESCS) index (Tables B7.1 and B7.2). Though the correlation with both scores is weak overall, it is nonetheless significant. Moreover, disadvantaged students – those in the bottom quartile of the PISA ESCS index within their country-language group sample – scored 0.75 points less than their advantaged peers (those in the within-group top quartile of the PISA ESCS index) on the standardised holistic creativity scale, on average, which represents half a standard deviation in performance (Table B7.3).

The mean response appropriateness, originality and value scores of advantaged and disadvantaged students reveal interesting variation in the sources of this socio-economic gap in creative thinking performance. Figure 7.1 shows that the largest score gaps occur with respect to mean response appropriateness and flexibility scores: advantaged students score 0.3 score points and 0.2 score points higher than their disadvantaged peers on these criteria, respectively, on average across country-language groups (Table B7.4). The mean score gap for response originality and value scores are also just below 0.2 score points, though these results are significantly influenced by the removal of non-appropriate ideas from the analyses (see Box 7.2). These differences in task scores between advantaged and disadvantaged students represent around one-third of a standard deviation in performance for appropriateness scores compared to around one-fifth of a standard deviation in performance for originality and value scores.

Box 7.1. Which score data are used in this chapter?

The analyses in this chapter draw upon two types of rescored data:

- Holistic creativity scores: Each student response was scored based on judges' evaluations of the overall creative quality of the response. The holistic creativity score reflects a relative ranking of creative quality compared to all other responses in that country-language-task group. To account for differences in the distribution of scores across country-language-task groups, the analyses in this chapter standardised and transformed the raw holistic creativity scores (original range 1-7 score points) onto a scale that ranges from 1-10 score points, with a mean of 5 score points.
- Criteria scores: Each idea within a student response was scored for its appropriateness, originality and value. A cumulative idea sum score was also produced by aggregating the scores across these three criteria. Tasks classified as 'generate diverse ideas' tasks required students to submit multiple ideas per response; for these tasks, criteria scores across ideas were averaged so that each student response is associated with a single appropriateness, originality and value score.

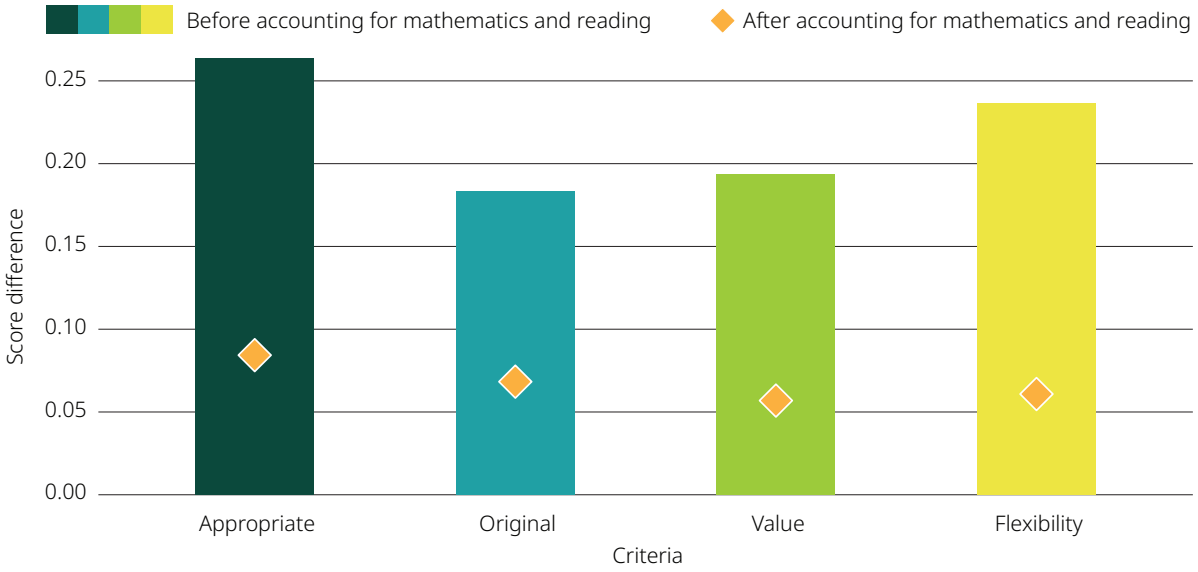
Terms used in this chapter

- "Advantaged students": Students that score in the top quartile of the PISA economic, social and cultural (ESCS) index, with respect to the country-language group sample included in the PISA CT Rescoring project.
- "Disadvantaged students": Students that score in the bottom quartile of the PISA economic, social and cultural (ESCS) index, with respect to the country-language group sample included in the PISA CT Rescoring project.

This chapter focuses on examining differences in idea quality between advantaged and disadvantaged student groups. For more information on the scoring methods, see Annex A2.

Figure 7.1. Disadvantaged students struggle the most in coming up with appropriate ideas

Difference in mean response criteria scores between advantaged and disadvantaged students (all tasks), on average across country-language groups, before and after accounting for mathematics and reading



Note: All differences are significant (bars and markers). Bars represent mean score differences between advantaged and disadvantaged students before accounting for mathematics and reading scores, markers represent mean score differences after accounting for these factors. Mean score differences were computed on the pooled responses across all tasks in each country-language group, then averaged across country-language groups.

Source: Table B7.4

Box 7.2. Treatment of non-appropriate ideas for some analyses in this chapter

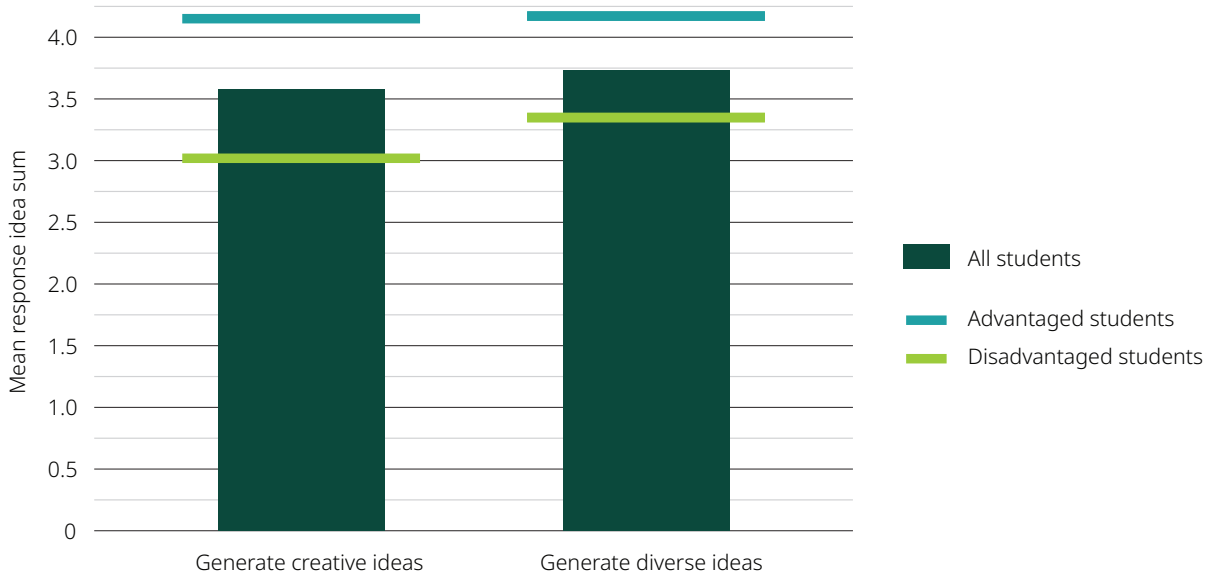
When calculating the mean response originality and value scores, only appropriate ideas were retained. According to the scoring methodology for the criteria-based approach (see Annex A2), judges were instructed to first score student ideas for appropriateness. Any ideas that achieved an appropriateness score of 0 were not scored further, meaning there are no originality or value scores associated with non-appropriate ideas. This process was followed to minimise the coding burden on judges and to simplify the interpretation of the criteria-based scores, following the argument that inappropriate ideas may well be original but they cannot be creative (Runco and Jaeger, 2012[16]).

Differences in mean response criteria scores between disadvantaged and advantaged students should be interpreted with this methodology in mind. For example, the mean appropriateness scores of advantaged and disadvantaged students are calculated using all idea scores, including those from students that scored 0 on idea appropriateness due to a lack of meaningful engagement with the task. In contrast, the mean score gap observed between advantaged and disadvantaged students for response originality and value, respectively, is calculated on only partially or fully appropriate ideas, and therefore likely excludes the most disengaged students from the sample in this analysis. As a result, it would be reasonable to expect the score difference between advantaged and disadvantaged students in mean response originality and value to be smaller than the equivalent in mean appropriateness. Indeed, when the originality and value scores of non-appropriate ideas are included in the analyses (recoded as 0 scores), then the score differences between advantaged and disadvantaged students are greater in mean response originality and value than in mean response appropriateness (Table B7.5).

The relatively large difference between advantaged and disadvantaged students in mean response appropriateness scores may be influenced by two factors. First, students from disadvantaged backgrounds may struggle more with the convergent thinking processes involved in creative thinking that support the evaluation and selection of ideas that are best suited for a given purpose. While the results from the PISA 2022 assessment found that the largest performance differences between advantaged and disadvantaged students were observed in the 'generate diverse ideas' tasks (OECD, 2024[1]), the PISA CT Rescoring study data suggest that disadvantaged students scored relatively worse in the 'generate creative ideas' tasks included in the study than they did in the 'generate diverse ideas' tasks (Figure 7.2). Our results in Chapter 5 of this report also broadly suggest that advantaged students may be more successful at generating creative ideas within more appropriate constraints than their disadvantaged peers. However, more research is needed to better understand potential differences in the cognitive processes involved in creative thinking between advantaged and disadvantaged students.

Figure 7.2. Disadvantaged students scored comparatively lower in 'generate creative ideas' tasks than in 'generate diverse ideas' tasks, relative to advantaged students

Mean response idea sum scores of advantaged and disadvantaged students, on average across country-language groups, by ideation process



Note: Average scores are computed by task first in each country-language group and then averaged across relevant tasks in each task group. The idea sum score refers to the sum of the idea appropriateness, originality and value scores. For 'generate diverse ideas' items that contain multiple ideas per response, the idea sum scores for all ideas were averaged to give a single mean idea sum score per student response.

Source: Table B2.2 and Table B7.6

The second factor that may play an even bigger influence in socio-economic differences in creative thinking performance is domain readiness. It stands to reason that proficiency in the core academic subjects provides students with a solid foundational knowledge and skills that are beneficial to creative thinking in a range of contexts, possibly explaining why socio-economic differences in creative thinking have been found to weaken with age and as students become more educated (Acar et al., 2023[2]). Indeed, students' mathematics and reading scores accounted for a large proportion of the observed differences in creative thinking performance by socio-economic status in the PISA 2022 results (OECD, 2024[1]).

Do reading and mathematics scores similarly account for most of the observed performance gap between advantaged and disadvantaged students in the PISA CT Rescoring project results? Differences in standardised holistic creativity scores between advantaged and disadvantaged students reduce from 0.75 score points to just under 0.3 score points (i.e. from half a standard deviation in performance to around one-fifth), after accounting for student mathematics and reading scores (Table B7.3). While this difference remains significant, on average across country-language groups, it notably becomes insignificant in Canada (English and French provinces), Chile, Germany, France and Saudi Arabia. Figure 7.1 also shows that the performance gap on each of the criteria-based scores is significantly reduced after accounting for student mathematics and reading scores: differences between advantaged and disadvantaged students in each of the score criteria fall to below 0.1 score points, on average across country-language groups. While these differences are small and often insignificant at the individual country-language group level, they remain significant on average (Table B7.4).

Overall, the much smaller gaps in creative thinking scores after accounting for students' reading and mathematics scores is cause for optimism, highlighting that a significant proportion of the socio-economic differences in creative thinking appear remediable with more equitable teaching and learning opportunities for students. However, significant differences still exist even after accounting, meaning that other factors beyond foundational literacies contribute to socio-economic disparities in creative thinking.

7.1.1. Are there cross-cultural differences in the relationship between socio-economic status and creative thinking scores?

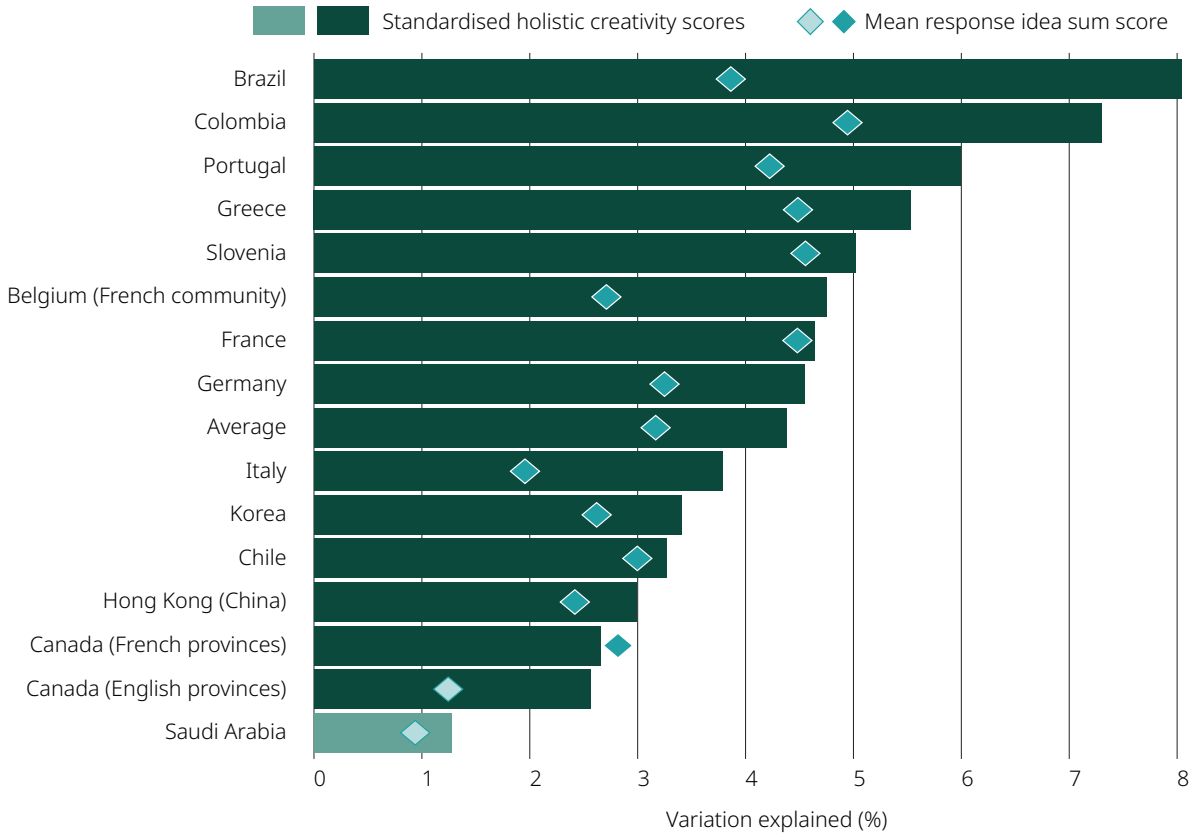
The association between socio-economic status and creative thinking scores varies across country-language groups. The standardised holistic creativity scores showed a stronger association with student socio-economic status across all country-language groups (Figure 7.3), perhaps because this scoring method captures or is influenced by broader aspects of student responses that might also be associated with socio-economic status (e.g. vocabulary, spelling and grammar, etc.). Additionally, the holistic judgement method evaluates student responses relative to their peers, and this forced comparison may contribute to the stronger association observed between holistic creativity scores and socio-economic status (i.e. regardless of large or small differences in overall idea quality, disadvantaged students tend to provide comparatively less creative answers than their advantaged peers).

The strongest associations between students' standardised holistic creativity scores and their socio-economic status were observed in Brazil, Colombia, Portugal and Greece, with over 5% of the variation in scores in Portugal and Greece and over 7% of the variation in scores in Colombia and Brazil explained by students' socio-economic index (Figure 7.3). In contrast, less than 3% of the variation in holistic creativity scores can be explained by socio-economic status in Hong Kong (China), Canada (English and French provinces) and Saudi Arabia. The criteria-based method follows a largely similar pattern, albeit the proportion of variation in students' mean response idea sum scores explained by socio-economic status is smaller in all country-language groups except for Canada (French provinces).

Despite differences across country-language groups, students with a higher socio-economic status score significantly higher than students with a lower socio-economic status, on average: a one-unit increase in the PISA ESCS index is associated with a difference of 0.36 standardised holistic creativity score points, which represents around a quarter of the standard deviation in performance amongst students. In Brazil and Colombia, the score difference is even larger – closer to a third of the standard deviation in performance. It is unsurprising that countries with greater within-country inequality in student socio-economic and cultural status provide relatively more educational opportunities to privileged students, which in turn may act to widen performance disparities in creative thinking. Indeed, Colombia, Brazil and Portugal all have significantly larger within-country variation in student socio-economic status than other country-language groups that participated in the PISA Rescoring project (OECD, 2023[17]).

Figure 7.3. Socio-economic status explains a larger proportion of the variation in holistic creativity scores than students' mean response idea sum scores

Proportion of variation in standardised holistic creativity scores (bar) and mean response idea sum scores (marker) explained by socio-economic status (all tasks), by country-language group



Note: Significant r-squared values (i.e. proportion of explained variation) are shown by bars/markers in a darker shade. Country-language groups are ordered in descending order of the proportion of explained variation in standardised holistic creativity scores.

Source: Table B7.8 and Table B7.9

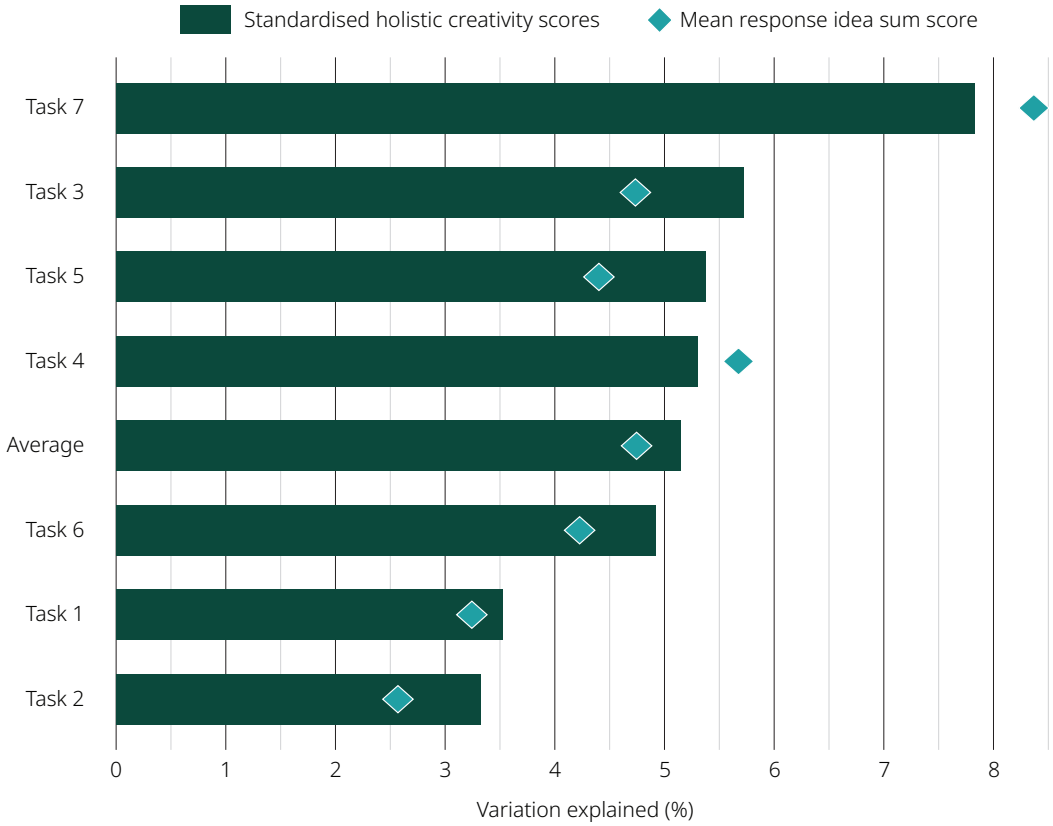
As discussed in the previous section, differences in the relationship between socio-economic status and creative thinking scores in each country-language group is also likely influenced by the equivalent relationship between academic achievement and socio-economic status more broadly (Dai et al., 2012[18]). Some countries and economies with a stronger association between socio-economic status and mathematics performance (OECD, 2023[17]), for example Belgium, France and Colombia, also show above-average variance in creative thinking scores explained by socio-economic status. After accounting for both mathematics and reading scores, the influence of socio-economic status on students' standardised holistic creativity scores reduces substantially but remains significant in all country-language groups except Chile, Germany, France and Saudi Arabia (Table B7.10).

7.2. Are socioeconomic differences more pronounced in some types of task?

The results so far have shown that the association between socio-economic status and creative thinking performance persists across country-language groups and scoring methods, and that disadvantaged students struggled relatively more in the 'generate creative ideas' tasks than in the 'generate diverse ideas' tasks included in the PISA CT Rescoring study (Figure 7.2). Looking at the scores of students on a task-by-task basis, Figure 7.4 shows that the variation in creativity scores explained by socio-economic status ranges across tasks – from around 3% of the variation in Task 2 (the *Illustration Titles* task; see Figure C in the Introduction) to close to 8% of the variation in Task 7 (*The Exhibit*; see Figure H).

Figure 7.4. Tasks differ in how much socio-economic status influences student creativity scores

Proportion of variation in standardised holistic creativity scores (bar) and mean response idea sum scores (marker) explained by socio-economic status, by task (on average across country-language groups)



Note: All r-squared values (i.e. proportion of variation explained) are significant. Tasks are ordered in descending order of the proportion of explained variation in standardised holistic creativity scores.

Source: Table B7.8 and Table B7.9.

There are mixed findings in previous research on the effects of task modality and socio-economic differences in performance in creative thinking tasks. Results from the PISA 2022 assessment, as well as other empirical studies, suggest that verbal tasks (i.e. tasks requiring writing) tend to exacerbate socio-economic differences in performance compared to those that require students to demonstrate creativity non-verbally (OECD, 2024[1]; Gajda, Karwowski and Beghetto, 2017[19]). However, a recent meta-analysis found that there were no significant differences in the performance of students with different socio-economic status on verbal and non-verbal creativity tasks (Acar et al., 2023[2]).

In the PISA CT Rescoring study results, when looking at the average character count of student responses to different tasks, there is an overlap between the average verbosity of tasks and the strength of the association between student scores and socio-economic status on the task (see Table 7.1). For example, the three tasks with the strongest association between student scores and socio-economic status place significant writing demands on students compared to the other tasks included in the study. In *The Exhibit* (Task 7) students are asked to describe three elaborate engineering solutions for a problem; the task *2983* (Task 3; see Figure D in the Introduction) asks students to describe an idea for a story; and the *Food Waste* task (Task 5; see Figure F) requires students to describe three solutions to address supermarket waste. Conversely, both the *Science Fair Poster* and *Illustration Titles* tasks (Tasks 1 and 2; see Figures B and C) require limited writing from students and show much weaker associations with socio-economic status and student scores on average across country-language groups.

Table 7.1. Some tasks have higher response demands than others

Mean response character count by task, on average across country-language groups

Task number	Task title	Mean response character count
Task 2	Illustration Titles	30
Task 1	Science Fair Poster	71
Task 4	Save the Bees	78
Task 5	Food Waste	85
Task 6	Save the River	101
Task 7	The Exhibit	110
Task 3	2983	203

Note: Tasks are ordered in ascending order of average character count (lowest to highest). In the *Science Fair Poster* task, the character count refers to the optional poster description response box that students can choose to complete to accompany their poster design. Mean response character counts are rounded to the nearest integer.

Source: Table B7.12

While mean character count alone does not necessarily equate to task demand, nor to the verbal proficiency required to respond sufficiently to a given task, being able to successfully convey one's ideas will inevitably affect the extent to which judges understand and value them – especially in tasks that require more elaborate responses. Although judges were explicitly instructed to focus on response content rather than grammatical form or style, it may be that judges were influenced by these factors or the level of detail provided within student responses. After accounting for students' mathematics and reading scores, the association between students' creativity scores and their socio-economic status considerably reduces in each task, though it remains significant across all tasks (Table B7.10).

Task context and instructions might also have a significant effect on how creativity tasks are approached by different student groups. For example, the limits of an appropriate response are likely more constrained within an engineering problem-solving task compared to a graphic design task or creative writing exercise, which may in turn create the impression or expectation that "correct" answers exist in some types of task. Indeed, the results in the previous section of this chapter show that appropriateness was the most challenging criterion for disadvantaged students to achieve, and the results in Chapter 1 show that response appropriateness tends to be more correlated with creative scores in the problem-solving tasks than the expressive tasks. Less confident students may therefore struggle to think of creative ideas in more practical problem-solving tasks for fear of not providing a "good enough" answer (see next section).

7.3. What factors might moderate the relationship between socio-economic status and creative thinking performance?

7.3.1. Self-efficacy beliefs

Some research has suggested that socio-economic differences in creative thinking performance may be mediated by the testing environment itself (Acar et al., 2023[2]). When a task is understood as a test, it can activate self-directed beliefs that shape performance (Spencer and Castano, 2007[20]). Students from disadvantaged socio-economic backgrounds are more likely to hold unfavourable beliefs towards their own creative abilities, leading to lower creative scores through less engagement and risk-taking (Yang et al., 2020[21]; Beghetto, Karwowski and Reiter-Palmon, 2021[22]; Puente-Díaz, 2016[23]).

Our data shows a positive but weak correlation between students' standardised holistic creativity scores and their creative self-efficacy and mathematics self-efficacy, respectively, on average across country-language groups (Table B7.13). Figure 7.5 shows that the gap in standardised holistic creativity scores between advantaged and disadvantaged students reduces after accounting for students' self-efficacy beliefs across all country-language groups, albeit to varying degrees. The largest reductions in score differences were observed in Belgium (French community) (0.35 score points), Greece and Germany (both 0.24 score points), in contrast to only small differences in the score gap observed in Hong Kong (China) and Colombia. Despite moderating the relationship between socio-economic status and creative thinking scores to some degree in each country-language group, belonging to the highest socio-economic quartile and not the lowest remains a significant predictor of success in creative thinking across the board, even after accounting for students' self-efficacy beliefs.

The role of self-efficacy beliefs in moderating socio-economic differences in performance also varies by task. The score difference (holistic creativity scores) between advantaged and disadvantaged students in *The Exhibit* and the *Save the River* tasks – the two scientific problem-solving tasks – reduces by close to 0.2 score points, after accounting for students' creative and mathematics self-efficacy beliefs (Table B7.15). Across the other tasks included in the study the relative score difference after accounting for self-efficacy beliefs also reduces, though to a lesser degree (between 0.12 and 0.16 score points).

These findings on self-efficacy, creative thinking scores and student background suggest that cultivating self-efficacy beliefs is important for supporting creative performance in all types of tasks. Helping students to feel confident in both their creative and general academic abilities may be especially important for supporting students from disadvantaged backgrounds in tasks that are more difficult and/or that have a more explicit focus on domain knowledge and skills (like scientific creativity).

Figure 7.5. Self-efficacy beliefs moderate socio-economic disparities in standardised holistic creativity scores in all country-language groups

Mean difference in standardised holistic creativity scores between advantaged and disadvantaged students (all tasks), before and after accounting for self-efficacy beliefs, by country-language group

Note: All differences (before and after accounting for self-efficacy beliefs) are significant. Country-language groups are ordered by the largest reduction in the difference in standardised holistic creativity scores between advantaged and disadvantaged students after accounting for student self-efficacy beliefs (creative self-efficacy and mathematics self-efficacy). Mean score differences between advantaged and disadvantaged students are computed on the pooled responses across tasks in each country-language group.

Source: Table B7.14

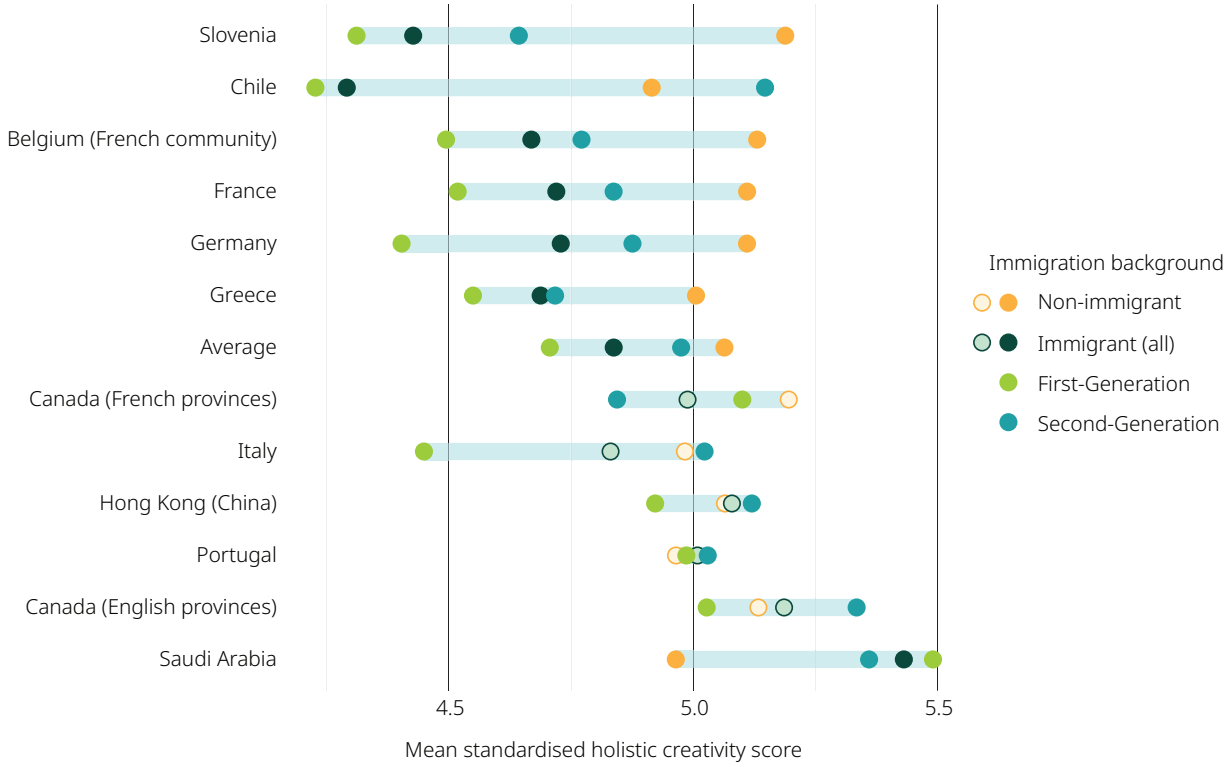
7.3.2. Diverse life experiences: Multilingualism and immigration status

As described in the introduction to this chapter, diverse life experiences may support creative thinking skills. While socio-economic advantage can facilitate access to a range of experiences and opportunities, not all diverse life experiences are inherently costly. Being fluent in multiple languages and immersion in different cultures, for example, have been independently shown to support creative potential (Fürst and Grin, 2021[24]).These experiences can enhance executive functions and cognitive flexibility, enabling individuals to recognise different perspectives and create meaningful associations between diverse concepts, which in turn strengthen fluency, elaboration, originality and flexibility in creative tasks (Ghonsooly and Showqi, 2012[25]; Tadmor, Tetlock and Peng, 2009[26]; Tadmor, Galinsky and Maddux, 2012[27]). While language fluency and exposure to different cultures might be developed through travel, it can also be facilitated through living in different countries and/or having family members from a different socio-cultural or linguistic background.

Findings from the PISA CT Rescoring study show that student immigrant status can have a mixed influence on creative thinking scores of students. Figure 7.6 shows students' mean standardised holistic creativity scores by immigrant background group (no immigrant background, first-generation, second-generation). On average across country-language groups, students with no immigrant background have a small but statistically significant advantage over their peers with an immigrant background (score difference of around 0.2 points, or less than a sixth of a standard deviation in scores). However, the size and direction of score differences between immigrant and non-immigrant background students vary across country-language groups: score differences in favour of non-immigrant background students were largest in Slovenia, Chile, Belgium (French community), France, Germany and Greece; statistically insignificant in Italy, Hong Kong (China), Portugal, Canada (English and French provinces); and reversed in Saudi Arabia (students with an immigrant background scored nearly 0.5 score points more than those with no immigrant background, or around one third of a standard deviation higher).

Figure 7.6. Students with an immigrant background tend to achieve lower scores than students without an immigrant background

Mean standardised holistic creativity scores by immigrant background status (all tasks), by country-language group



Note: Statistically significant differences between immigrant (all) and non-immigrant students are shown by filled markers. Country-language groups are ordered by the mean difference in standardised holistic creativity scores between immigrant (all) and non-immigrant students (non-immigrant - immigrant) (highest to lowest). Country-language groups with fewer than 20 first- or second-generation immigrant students in the sample were excluded from the analysis. Mean score differences are computed on the pooled responses across tasks in each country-language group.

Source: Table B7.16

Differences in creativity scores between students with an immigrant background and those without an immigrant background seem to be largely influenced by differences in response appropriateness scores, which in turn appear driven by mathematics and reading scores. Figure 7.7 shows that in six of the 12 country-language groups included in the analysis, the largest difference in response quality between immigrant and non-immigrant-background students relates to response appropriateness scores, with students with an immigrant background scoring around 0.1 score points lower on average across country-language groups than students with no immigrant background.

Figure 7.7. Mathematics and reading scores explain most of the differences in scores by immigrant background

Differences in mean response criteria score (appropriate, original and value) between non-immigrant and immigrant background students (all tasks), before and after accounting for mathematics and reading scores

Note: Significant differences are shown by colours in a darker shade for each criteria (bars) before accounting, and by markers of a darker shade after accounting. Mean score differences were computed on the pooled responses across all tasks in each country-language group.

Source: Table B7.17

While these results might be driven in part by the treatment of non-appropriate responses when comparing originality and value scores between student groups (see Box 7.2), they also appear closely related to differences in students' reading and mathematics proficiency. After accounting for student scores in the core domains, immigrant students in many country-language groups tended to perform no differently to their peers without an immigrant background across the different score criteria. On average, and in Belgium (French community), Canada (French provinces), France and Slovenia, the response appropriateness scores of students with an immigrant background even reversed from being significantly lower than their non-immigrant background peers to significantly higher than their peers, after accounting for reading and mathematics scores. The same reversal after accounting for mathematics and reading was also observed in Greece for response value scores and Slovenia for response originality scores, with immigrant background students significantly outperforming their non-immigrant counterparts, and in Saudi Arabia, where non-immigrant background students then significantly outperform their immigrant background peers in response originality and value scores.

Overall, after accounting for students' reading and mathematics scores, students with an immigrant background tend to score slightly but significantly higher than students without an immigrant background – both in their standardised holistic creativity scores and in their mean response idea sum scores. Taken all together, these findings suggest that students with an immigrant background tend to score lower on the creative thinking tasks not because of their immigrant background *per se*, but perhaps rather due to relative difficulties in understanding and expressing themselves in the test language. When students with similar reading proficiency are compared, immigrant students perform no worse than non-immigrant students in response originality and value scores, and even significantly outperform their non-immigrant peers in response appropriateness and in overall response quality (standardised holistic creativity scores and mean response idea sum scores), on average.

Language proficiency in multiple languages is also associated with higher creativity scores across the two scoring methods used in this chapter (Table B7.17). Students that can speak two or more languages well enough to converse with others scored significantly higher than students that speak only one language (a difference of 0.2 score points on the standardised holistic creativity scale and mean response idea sum scores, respectively), on average across country-language groups. However, after accounting for mathematics and reading scores, there are no significant differences in mean response idea sum score between students that can speak more than one language and those that cannot, and students that speak only one language in fact achieve higher standardised creativity scores (difference of 0.1 score points) after accounting for these factors. It may be that students who are more likely to be able to converse with others in multiple languages are those students that perform better academically, in general, and this largely accounts for the significant differences in performance observed before accounting for reading and mathematics performance.

In sum, the findings are mixed when it comes to the relationship between diverse life experiences and their relationship to creative thinking. While the results suggest that factors like immigration background and multi-linguicism do not adversely affect creative thinking, they also find that most of the difference in performance between student groups (positive or negative) is better explained by differences in reading and mathematics scores. Of course, it may also be that immigrant background and language proficiency, as measured in PISA (i.e. student questionnaire self-reports), are not the best proxies of diverse life experience, and/or that other diverse life experiences and opportunities might better mediate differences in creative thinking performance.

Key findings and implications

- Student socio-economic status and scores on the creative thinking tasks are significantly positively associated, in general, though the strength of the relationship is weak (r=0.2). However, a one-unit increase in the PISA ESCS index is associated with a significant increase in standardised holistic creativity score of 0.36 points (around a quarter of a standard deviation in performance).
  - The association between socio-economic status and creative thinking performance was strongest in countries with the largest within-country inequality (e.g. Colombia, Brazil and Portugal). It may be that privileged students in these countries have relatively better access to quality educational opportunities than their disadvantaged peers, which in turn widens performance disparities.
- Disadvantaged students scored 0.75 points less than advantaged students (around half a standard deviation in holistic creativity scores), on average. Disadvantaged students also scored lower in each of the criteria-based scores than advantaged students, particularly response appropriateness and flexibility.
- Differences in reading and mathematics proficiency between advantaged and disadvantaged student groups largely account for their performance differences in creative thinking.
  - Policy efforts should focus on providing more equitable teaching and learning experiences, and opportunities to practice creative thinking, for all students.
  - However, not all of the difference in creative thinking scores can be explained by academic performance – meaning that other factors beyond foundational literacies also contribute to socio-economic differences in performance.
- Disadvantaged students struggled more in certain types of creative thinking tasks, in particular convergent thinking tasks, tasks with greater writing demands, and problem-solving tasks.
  - Students from disadvantaged backgrounds may struggle with the convergent thinking processes involved in creative thinking that support the evaluation and selection of ideas that are best suited for a given purpose.
  - Disadvantaged students may also struggle to successfully convey their ideas, in part given their lower proficiency in core literacies (on average), or due to greater disengagement with tasks with higher responses demands.
  - Given differences in the scores of students in the PISA core domains of reading and mathematics, disadvantaged students may have less "domain readiness" than advantaged students, i.e. relevant knowledge and skills related to a given problem context. Moreover, problem-solving tasks tend also to be more constrained in terms of appropriateness than expressive tasks, meaning it is relatively harder to come up with original and valuable ideas that are also task appropriate.
- Self-efficacy beliefs effectively moderate students' success in creative thinking tasks in all country-language groups, and especially in scientific problem-solving tasks (on average across country-language groups).

References

Acar, S. et al. (2023), "Socio-economic status and creativity: A meta-analysis", *The Journal of Creative Behavior*, Vol. 57/1, pp. 138-172, https://doi.org/10.1002/jocb.568. [2]

Batey, M. and A. Furnham (2006), "Creativity, intelligence, and personality: A critical review of the scattered literature", *Genetic, Social and General Psychology Monographs*, Vol. 132/4, pp. 355-429. [8]

Beghetto, R., M. Karwowski and R. Reiter-Palmon (2021), "Intellectual risk taking: A moderating link between creative confidence and creative behavior?", *Psychology of Aesthetics, Creativity, and the Arts,* Vol. 15/4, pp. 637-644, https://doi.org/10.1037/aca0000323. [22]

Castillo-Vergara, M. et al. (2018), "Does socioeconomic status influence student creativity?", *Thinking Skills and Creativity,* Vol. 29, pp. 142-152, https://doi.org/10.1016/j.tsc.2018.07.005. [3]

Dai, D. et al. (2012), "Influences of social and educational environments on creativity during adolescence: Does SES matter?", *Creativity Research Journal,* Vol. 24/2-3, pp. 191-199, https://doi.org/10.1080/10400419.2012.677338. [18]

Ellis, B. et al. (2023), *The hidden talents framework: Implications for science, policy and practice,* Cambridge Univeristy Press, https://doi.org/10.1017/9781009350051. [10]

Fürst, G. and F. Grin (2021), "Multicultural experience and multilingualism as predictors of creativity", *International Journal of Bilingualism,* Vol. 25/5, pp. 1486-1494, https://doi.org/10.1177/13670069211019468. [24]

Gajda, A., M. Karwowski and R. Beghetto (2017), "Creativity and academic achievement: A meta-analysis", *Journal of Educational Psychology,* Vol. 109/2, pp. 269-299, https://doi.org/10.1037/edu0000133. [19]

Ghonsooly, B. and S. Showqi (2012), "The effects of foreign language learning on creativity", *English Language Teaching*, Vol. 5/4, pp. 161-167, https://doi.org/10.5539/elt.v5n4p161. [25]

Gocłowska, M. and R. Crisp (2014), "How dual-identity processes foster creativity", *Review of General Psychology,* Vol. 18/3, pp. 216-236, http://dx.doi.org/10.1037/gpr0000008216. [6]

Jankowska, D., I. Lebuda and J. Gralewski (2024), "Creating home: Socioeconomic status and home environment as predictors of family climate for creativity", *Thinking Skills and Creativity,* Vol. 52, pp. 1-13, https://doi.org/10.1016/j.tsc.2024.101511. [5]

Karwowski, M. (2011), "It doesn't hurt to ask… but sometimes it hurts to believe: Polish students' creative self-efficacy and its predictors", *Psychology of Aesthetics, Creativity, and the Arts,* Vol. 5/2, pp. 154-164, https://doi.org/10.1037/A0021427. [14]

Kupczyszyn, K., V. Filippetti and L. Oros (2024), "Socioeconomic status effects on children's creativity: The mediating role of executive functions", *Thinking Skills and Creativity,* Vol. 51, pp. 1-15, https://doi.org/10.1016/j.tsc.2023.101437. [15]

Lee, H. and K. Kim (2011), "Can speaking more languages enhance your creativity? Relationship between bilingualism and creative potential among Korean American students with multicultural link", *Personality and Individual Differences*, Vol. 50/8, pp. 1186-1190, https://doi.org/10.1016/j.paid.2011.01.039. [12]

Lu, Z., Y. Ding and Y. Nie (2024), "How does family socioeconomic status affect creativity? The role of creative self-efficacy and critical thinking disposition", *Current Psychology,* Vol. 43/6, pp. 5674-5681, https://doi.org/10.1007/s12144-023-04768-x. [13]

OECD (2024), *PISA 2022 Results (Volume III): Creative Minds, Creative Schools,* PISA, OECD Publishing, https://doi.org/10.1787/765ee8c2-en. [1]

OECD (2023), *PISA 2022 Results (Volume I): The State of Learning and Equity in Education,* OECD Publishing, Paris, https://doi.org/10.1787/53f23881-en. [17]

Puente-Díaz, R. (2016), "Creative self-efficacy: An exploration of its antecedents, consequences, and applied Implications", *The Journal of Psychology,* Vol. 150/2, pp. 175-195, https://doi.org/10.1080/00223980.2015.1051498. [23]

Ritter, S. et al. (2012), "Diversifying experiences enhance cognitive flexibility", *Journal of Experimental Social Psychology,* Vol. 48/4, pp. 961-964, https://doi.org/10.1016/j.jesp.2012.02.009. [7]

Runco, M. and J. Jaeger (2012), "The standard definition of creativity", *Creativity Research Journal*, Vol. 24/1, pp. 92-96, https://doi.org/10.1080/10400419.2012.650092. [16]

Spencer, B. and E. Castano (2007), "Social class is dead. Long live social class! Stereotype threat among low socioeconomic status individuals", *Social Justice Research*, Vol. 20/4, pp. 418-432, https://doi.org/10.1007/s11211-007-0047-7. [20]

Tadmor, C., A. Galinsky and W. Maddux (2012), "Getting the most out of living abroad: Biculturalism and integrative complexity as key drivers of creative and professional success", *Journal of Personality and Social Psychology*, Vol. 103/3, pp. 520-542, https://doi.org/10.1037/A0029360. [27]

Tadmor, C., P. Tetlock and K. Peng (2009), "Acculturation strategies and integrative complexity: The cognitive implications of biculturalism", *Journal of Cross-Cultural Psychology,* Vol. 40/1, pp. 105-139, https://doi.org/10.1177/0022022108326279. [26]

Xu, X. and W. Pang (2020), "Reading thousands of books and traveling thousands of miles: Diversity of life experience mediates the relationship between family SES and creativity", *Scandinavian Journal of Psychology,* Vol. 61/2, pp. 177-182, https://doi.org/10.1111/sjop.12591. [4]

Yang, Y. et al. (2020), "Hope and creative self-efficacy as sequential mediators in the relationship between family socioeconomic status and creativity", *Frontiers in Psychology*, Vol. 11/438, pp. 1-8, https://doi.org/10.3389/fpsyg.2020.00438. [21]

Young, E. et al. (2022), "Hidden talents in context: Cognitive performance with abstract versus ecological stimuli among adversity-exposed youth", *Child Development,* Vol. 93/5, pp. 1493-1510, https://doi.org/10.1111/cdev.13766. [11]

Zhang, D. et al. (2018), "Family socio-economic status and parent-child relationships are associated with the social creativity of elementary school children: The mediating role of personality traits", *Journal of Child and Family Studies*, Vol. 27/9, pp. 2999-3007, https://doi.org/10.1007/s10826-018-1130-4. [9]

Annexes

- **Annex A1**: Sample sizes and sampling methodology in the PISA CT Rescoring project
- **Annex A2**: Scoring methods and coding design in the PISA CT Rescoring project
- **Annex B**: Figures and tables available online
- **Annex C1**: Full C-K schema for Food Waste task
- **Annex C2**: Comprehensive node category list for Food Waste task
- **Annex C3**: Full C-K schema for The Exhibit task
- **Annex C4**: Comprehensive node category list for The Exhibit task

Annex A1.
Sample sizes and sampling methodology in the PISA CT Rescoring project

This annex is available online at:
https://www.oecd.org/en/publications/seven-questions-about-creativity-and-creative-thinking_0aa52128-en/support-materials.html.

Annex A2.
Scoring methods and coding design in the PISA CT Rescoring project

Introduction

The PISA 2022 Creative Thinking assessment framework describes the scoring rules applied in the PISA 2022 test (OECD, 2023[1]). The international scoring rules focused on evaluating students' capacity to generate "sufficiently" original or different ideas (depending on the item type) across task contexts, and accordingly, student responses were awarded either full credit, partial credit (in some cases), or no credit. This approach was adopted to maximise the scalability and reliability of the human coding effort in each country/economy and to facilitate the cross-cultural comparability of scores by focusing on scoring criteria that could be evaluated most objectively across diverse contexts (Foster and Piacentini, 2025[2]). However, the trade-off of the PISA 2022 scoring approach was that it could not identify the most creative responses within each participating country/economy, nor did it evaluate the overall "creative value" of students' responses (Foster and Piacentini, forthcoming[3]).

The three scoring methods applied in the PISA CT Rescoring project conceptualise and prioritise creative value in different ways to each other and to the international scoring rules applied in the PISA 2022 Creative Thinking assessment. The three scoring methods applied to the PISA CT Rescoring project data sample were:

- The **Holistic Judgement method**: Student responses were given a single "creativity" score reflecting the overall creative quality of the response, considering all relevant elements together.
- The **Criteria-Based method**: Each idea within a student response was scored for its appropriateness, originality and value, and responses to tasks requiring multiple ideas were scored for idea flexibility.
- The **Conceptual Knowledge Tree method**: Each idea within a student response was mapped onto a branch and node of a conceptual knowledge tree that defines all possible solution path ideas in a problem space.

The Holistic Judgement method and Criteria-Based method were applied to all student responses in the PISA CT Rescoring project sample. The Conceptual Knowledge Tree method was applied to two problem-solving tasks only, Task 5 – *Food Waste* and Task 7 – *The Exhibit* (see Figures F and H in the Introduction).

Holistic Judgement Method

Objective of scoring method

The Holistic Judgement scoring method involves attributing valid student responses with a single score between 1 (least creative) and 7 (most creative), based on judges' holistic judgement of the creativity of the response. Scores should reflect judges' subjective appraisal of the response, taking into account all aspects of its appropriateness, originality and usefulness, as well as idea flexibility in responses that include multiple ideas per response. The main objective of the holistic judgement method is to establish a relative ordering of the overall creative quality of responses within each country-language-task sub-sample, and to be able to identify the most creative responses within each national-language context.

Common rubric for attributing scores

Judges attributed scores with respect to within-country comparisons. Therefore, a holistic creativity score of 7 denotes that a response is considered to be amongst the most creative observed by the judge compared to all other responses in the country-language-task dataset. Likewise, a score of 4 denotes a response is considered of average quality with respect to the country-language-task sample, and a score of 1 denotes that a response is amongst the least creative observed within the country-language-task dataset.

To help judges interpret holistic creativity scores in a consistent way across country-language groups (i.e. to minimise culturally sensitive grading styles), a high-level rubric was provided that divided the available holistic creativity scores into three "score bands" and that described, in general terms, the overall quality of responses corresponding to each score band.

Judges were instructed to use the rubric as a guideline rather than a strict prescription, given that scores attributed according to this scoring method are ultimately subjective in nature. Judges were also instructed to approximate a normal distribution in their holistic creativity scores to facilitate relative ranking comparisons across country-language groups.

Score band 0 points (non-appropriate or missing responses)

Responses that are non-appropriate (i.e. are not meaningful or do not clearly connect to the task prompt) or that are missing should be coded as 0 score points.

Score band 1-2 points ("*low creativity*" responses)

Responses in this score band may reflect a limited effort to engage seriously with the task, may not be wholly relevant or appropriate considering the task constraints, and may lack original, thoughtful, interesting or noteworthy elaboration. Responses show minimal imagination and are conventional in their approach, with little personal flair.

- Around 10-20% of all responses in the country-language-task sample should be scored within this band.
- Note: For 'generate diverse ideas' tasks that require multiple ideas per student response: if students provide only one appropriate solution idea, they should be awarded a maximum score of 2 points.

Score band 3-5 points ("*average creativity*" responses)

Responses in this score band may reflect a reasonable level of engagement with the task but tend to rely on relatively common or obvious idea associations that lack further original, thoughtful, interesting or noteworthy elaboration. Responses show some signs of imagination but in a relatively conventional approach to the task that respects expected rules and conventions, albeit with some personal flair.

- Around 60-80% of all responses in the country-language-task sample should be scored within this band.
- Note: For 'generate diverse ideas' tasks that require three ideas per student response: if students provide only two appropriate solution ideas, they should be awarded a maximum score of 5 points.

Score band 6-7 points ("*high creativity*" responses)

Responses in this score band may reflect a significant engagement with the task, may propose relatively novel or original idea associations, and may add significant elaboration or detail that has some interesting, emotive, reflective, humorous or otherwise noteworthy impact on the judge. The response follows an innovative or original approach and demonstrates attention to detail. It effectively blends personal flair with technical knowledge.

- Around 10-20% of all responses in the country-language-task sample should be scored within this band.

Coding process

Coding design

For each task, at least two judges per country-language group were involved in scoring student responses so that inter-rater reliability metrics could be calculated and verified (see 'Inter-rater reliability' sub-section for more information). Multiple judges scored a sub-sample of around 100 responses per task, but not all student responses for each task were coded by multiple judges in all country-language groups.

In most country-language groups, a designated lead judge scored around 100 responses in the task sub-sample together with a second judge (i.e. the double coded sub-sample). The lead judge was responsible for verifying inter-rater reliability and implementing recalibration (if needed). The second judge scored the entirety of the task response sample. In Brazil and Portugal, coding teams implemented a balanced Incomplete Block Design, where each response in the task sample was coded by rotating pairs of judges. In this coding design, all possible combinations of two judges are represented in the coding of each task, with judges scoring batches of around 50 responses per pair of judges. A designated lead coder in each team was still responsible for verifying inter-rater reliability across judge pairs and implementing recalibration (if needed).

Coding steps

The Holistic Judgement method was the first of the three scoring methods applied by country-language teams in the PISA CT Rescoring project. Judges were instructed to follow the steps below when scoring responses according to the Holistic Judgement method:

1. As a first step, both judges assigned to a given task should familiarise themselves with all the responses in their country-language-task sample as well as the key pointers for scoring described in the PISA CT Rescoring project coding guide. As part of this familiarisation process, judges should identify and code any clearly aberrant, off-topic or missing responses (code 0).
2. Next, the designated lead judge should code 100 responses in the task sample (i.e. responses 1 to 100). The designated second judge should also code the same 100 responses in the task sample in reverse order (i.e. responses 100 to 1).
3. The two judges should then review their inter-rater reliability metrics for the double-coded sample of 100 responses and discuss their scores if flagged as requiring calibration attention (i.e. where one or more of the inter-reliability metric thresholds is exceeded). Judges should adjust their scores accordingly following calibration discussions until the inter-rater reliability metrics are within the acceptable range.
4. After this review (and where needed, calibration exercise), the second judge should proceed to score the entire response sample for the task.

Inter-rater reliability metrics

For the Holistic Judgement method, the main metrics used to evaluate inter-rater reliability across the double-coded sample were:

- Correlation coefficient: $r > 0.75$
- Mean score difference between judges: $\bar{x}$ difference < 1

If these thresholds were breached after completing the double-coded sample, then judges were instructed to meet to discuss and recalibrate their scores to try and reach, or at least approximate, the acceptable range(s). Across all tasks, all coding teams maintained a mean score difference between judges of less than one score point. In Brazil, the average correlation coefficient between judge pairs was below 0.75 in five tasks (Tasks 2, 4, 5, 6 and 7), although in three of these tasks (Tasks 5-7) the average correlation coefficient remained above 0.7.

Criterion-Based Method

Objective of scoring method

The Criteria-Based scoring method involves attributing student ideas within a response separate scores based on a series of individual criteria: appropriateness, originality, and value; and, where multiple ideas are involved, idea flexibility (i.e. for 'generate diverse ideas' tasks). In this scoring method, each idea within a student response receives multiple scores. The main objective of the Criteria-Based scoring method is to facilitate inter-rater agreement based on a set of common criteria recognised by the literature on creativity measurement as being important for defining creative ideas. A secondary objective is to identify cross-cultural differences in the relative weight attributed to these different criteria in the holistic appraisal of students' responses.

General description of criteria

To help judges understand and evaluate the scoring criteria in a consistent way across country-language groups, and to help judges assign scores consistently (i.e. to minimise culturally sensitive grading styles), country-language teams were provided with a high-level description of the scoring criteria (see below) that described the available score points for each criterion. In addition, teams were also provided with an item-specific contextualisation of the scoring criteria for each task. The item-specific criteria were developed and standardised based on the inputs of the country-language teams participating in the PISA CT Rescoring project.

Appropriateness (0-2 score points)

Appropriateness refers to the extent to which a response respects the task instructions and constraints, is in the required format, and is relevant to the task content. Scores for appropriateness range from 0 points (not appropriate) to 2 points (high appropriateness).

- Note: Ideas awarded a score of 0 points for appropriateness should not be further scored for originality, value or (where relevant) flexibility.

Originality (0-3 score points)

Originality refers to the extent to which a response presents a relatively uncommon, unusual, different, imaginative or innovative idea, or includes novel or imaginative details, with respect to other responses in the country-language sample. Scores for originality range from 0 points (not at all original) to 3 points (very original).

Value (0-3 score points)

Value refers to the extent to which a response is useful and impactful for its stated purpose. Scores for value range from 0 points (no value) to 3 points (very high value).

Flexibility (variable score points depending on number of ideas required)

Flexibility refers to the extent to which ideas within the same student response are different from each other. Scores for flexibility range from 0 points (not different) to 1 points (both ideas different), or 0 points (not different) to 2 points (all three ideas different), depending on the number of ideas required by the task.

- Note: for 'generate diverse ideas' tasks, all ideas within the response should be considered together when attributing a response flexibility score. There is therefore only one flexibility score per response.

Coding process

Coding design

For each task, at least two judges per country-language group were involved in scoring student responses so that inter-rater reliability metrics could be calculated and verified (see 'Inter-rater reliability' sub-section for more information). Multiple judges scored a sub-sample of around 100 responses per task, but not all student responses for each task were coded by multiple judges in all country-language groups.

In most country-language groups, a designated lead judge scored around 100 responses in the task sub-sample together with a second judge (i.e. the double coded sub-sample). The lead judge was responsible for verifying inter-rater reliability and implementing recalibration (if needed). The second judge scored the entirety of the task response sample. In Brazil and Portugal, coding teams implemented a balanced Incomplete Block Design, where each response in the task sample was coded by rotating pairs of judges. In this coding design, all possible combinations of two judges are represented in the coding of each task, with judges scoring batches of around 50 responses per pair of judges. A designated lead coder in each team was still responsible for verifying inter-rater reliability across judge pairs and implementing recalibration (if needed).

Coding steps

The Criteria-Based scoring method was applied after the Holistic Judgement scoring method. Judges were instructed to follow the steps below when scoring responses according to the Criteria-Based method:

1. As a first step, both judges assigned to a task should (re)-familiarise themselves with all the responses in their country-language-task sample as well as the key pointers for scoring described in the PISA CT Rescoring project coding guide.
2. Next, the designated lead judge should code 100 responses in the task sample (i.e. responses 1 to 100) according to the item-specific appropriateness criteria (score code 0, 1 or 2). The second judge should also code the same 100 responses in the sample according to the item-specific appropriateness criteria in reverse order to the lead judge (i.e. responses 100 to 1).
3. Then, the two judges should review their inter-rater reliability metrics for appropriateness for the double-coded sample of 100 responses and discuss their scores if flagged as requiring calibration attention (i.e. where one or more of the inter-reliability metric thresholds has been exceeded). Judges should adjust their scores accordingly following any calibration discussions until the inter-rater reliability metrics are within the acceptable range.
   - Note that any ideas that receive a score code of 0 for appropriateness should subsequently be treated as missing for the rest of the scoring exercise and should therefore not be scored for originality and value. Ideas that receive a score code of 0 for appropriateness should also not contribute towards the response flexibility score.
4. After this review (and where needed, calibration exercise), the lead judge should code the same 100 responses in the task sample in reverse order to Step 2 (i.e. responses 100 to 1) according to the item-specific originality criteria (score code 0, 1, 2 or 3). The second judge should also code the same 100 responses in the task sample in reverse order (i.e. responses 1 to 100).
5. Next, the lead judge and second judge should code the same 100 responses in the task sample in reverse order again (i.e. responses 1 to 100 for the lead judge, or responses 100 to 1 for the second judge) according to the item-specific value criteria (score code 0, 1, 2 or 3).
6. Finally, the lead judge and second judge should code the same 100 responses in the task sample in reverse order to the previous step according to the item-specific flexibility criteria (score code 0 or 1, or score code 0, 1 or 2, depending on the number of ideas required by the task).
   - Note that appropriate ideas in a student response should contribute to its flexibility score.
7. The two judges should then review their inter-rater reliability metrics for the double-coded sample of 100 responses for originality, value and flexibility scores. Judges should discuss their scores if flagged as requiring calibration attention (i.e. where one or more of the inter-reliability metric thresholds has been exceeded) for any of the score criteria. Judges should adjust their scores accordingly following any calibration discussions until the inter-rater reliability metrics are within the acceptable range for each score criteria.
8. After this review (and where needed, calibration exercise), the second judge should proceed to score the entire task sample for idea appropriateness, followed by idea originality, followed by idea value, and finally response flexibility. The second judge should continue to score all responses for each criterion before moving onto the next criterion.

Inter-rater reliability metrics

For the Criteria-Based scoring method, the main metrics used to evaluate inter-rater reliability across the double-coded sample for the four criteria were the same as for the Holistic Judgement method:

- Correlation coefficient: $r > 0.75$
- Mean score difference between judges: $\bar{x}$ difference < 1

If these thresholds were breached after completing the double-coded sample, then judges were instructed to meet to discuss and recalibrate their scores to try and reach, or at least approximate, the acceptable range(s). Across all tasks, all coding teams maintained a mean score difference between judges of less than one score point in all four of the score criteria. However, some teams did not meet the targeted correlation coefficient for all criteria in all tasks: in Brazil, the average correlation coefficient between judge pairs was below 0.75 for appropriateness in three tasks, for both originality and value in five tasks, and flexibility in three tasks; in Portugal, the average correlation coefficient between judge pairs was below the target threshold for appropriateness and originality in four tasks, and for value and flexibility in three tasks; in Belgium (French community), the correlation coefficient was not achieved for appropriateness and flexibility in one task, and for originality and value in two tasks; and in Hong Kong (China), the correlation coefficient was not achieved for value in one task.

Conceptual Knowledge (C-K) Tree Method

Objective of scoring method

The Conceptual-Knowledge (C-K) Tree scoring method has emerged in recent years for scoring problem-solving items in divergent thinking tasks (Hatchuel and Weil, 2009[4]; Agogué et al., 2014[5]). This approach makes it possible to generate, before obtaining any response data, an exhaustive mapping of relevant concepts and knowledge within a given problem and solution space. This objective mapping, often conceptualized as a "tree" with branches, sub-branches and nodes, allows judges to associate ideas within a solution response onto a well-defined schema from which ideational pathways can be inferred, as well as more granular idea flexibility and originality metrics.

Coding materials

To code responses according to this method, judges use two coding materials:

- C-K Tree schema: The C-K Tree schema represents all possible solution paths and is presented in the form of a decision tree, the ends of whose branches represent major categories of solutions (i.e. all solutions within a given sub-branch). Each major category (sub-branch) is labelled with a given letter (e.g. "A", "B", "C", etc.)
- List of nodes: Each sub-branch (i.e. major category) contains several nodes, which represent different, detailed types of responses. Each idea within a student response can be assigned to a particular node. Each node therefore has its own code, with the first letter of the code corresponding to the relevant sub-branch (e.g. for nodes within sub-branch "A", available node codes may be "Aa", "Ab", "Ac", and so on).

The full C-K Tree schema and comprehensive lists of nodes for Task 5 – *Food Waste* and Task 7 – *The Exhibit* can be found in Annexes C1-C4.

Coding process

The C-K Tree method was the last of the three scoring methods applied by country-language teams in the PISA CT Rescoring project.

Like the Holistic Judgement method and the Criteria-Based scoring method, two judges per country-language group were involved in scoring student responses so that inter-rater reliability metrics could be calculated and verified (see 'Inter-rater reliability' sub-section for more information). Multiple judges scored a sub-sample of at least 100 responses per task, but not all student responses for each task were coded by multiple judges in all country-language groups.

For each idea within a student response, judges were asked to assign ideas to the node of the C-K Tree schema that best represented the idea. In cases where an idea could be assigned to multiple nodes, judges could assign multiple codes (though the node that most closely maps to the idea should be assigned in the "main code" space).

Judges were instructed to follow the steps below when scoring responses according to the C-K Tree method:

1. As a first step, both judges assigned to a given task should familiarise themselves with the C-K Tree schema and the different major categories of responses. Note that in this scoring method, all ideas within a student response should be mapped onto a node of the C-K Tree schema, including ideas that might otherwise be considered non-appropriate (i.e. ideas that scored 0 points in the Holistic Judgement scoring method or that score 0 points for appropriateness in the Criteria-Based method).
2. Next, the designated lead judge should code 100 responses in the task sample (i.e. responses 1 to 100). The designated second judge should also code the same 100 responses in the task sample in reverse order (i.e. responses 100 to 1).
3. The two judges should then review their inter-rater reliability metrics for the double-coded sample of 100 responses and discuss their scores if flagged as requiring calibration attention (i.e. where one or more of the inter-reliability metric thresholds is exceeded). Judges should adjust their scores accordingly following calibration discussions until the inter-rater reliability metrics are within the acceptable range.
4. After this review (and where needed, calibration exercise), the second judge should proceed to score the entire response sample for the task.

Inter-rater reliability metrics

For the C-K Tree scoring method, the main metrics used to assess the inter-rater reliability of main code attributions across the double-coded sample were:

- Cohen's Kappa (k) coefficient: $k > 0.7$
- Number of times judges disagreed on the idea main code: $n < 60$ (or equivalent proportion of the double-coded sample, i.e. one-fifth of double-coded ideas).

If these thresholds were breached after completing the double-coded sample, then judges were instructed to meet to discuss and recalibrate their scores to try and reach, or at least approximate, the acceptable range(s). Across both tasks, all coding teams achieved a Cohen's kappa coefficient of above 0.7 and fewer than 60 (or equivalent proportion) of disagreements on the main code attribution of ideas. Some of the data for Colombia were recoded for both tasks to facilitate cross-cultural comparisons in the relative distribution of the data, though note that both pre- and post-recoding of the data the inter-rater reliability metrics were achieved.

References

Agogué, M. et al. (2014), "The impact of type of examples on originality: Explaining fixation and stimulation effects", *The Journal of Creative Behavior,* Vol. 48/1, pp. 1-12. [5]

Foster, N. and M. Piacentini (2025), "Assessing creative thinking at scale and across culture", in Katz-Buonincontro, J. and T. Kettler (eds.), *Oxford Handbook of Creativity and Education,* Oxford Handbooks, Oxford University Press. [2]

Foster, N. and M. Piacentini (forthcoming), "PISA 2022 creative thinking: What it is, what it is not, and what next?", *The Journal of Creative Behavior*, https://doi.org/10.1002/jocb.70068. [3]

Hatchuel, A. and B. Weil (2009), "CK design theory: An advanced formulation", *Research in Engineering Design,* Vol. 19, pp. 181-192. [4]

OECD (2023), *PISA 2022 Creative Thinking Framework,* PISA, OECD Publishing, https://doi.org/10.1787/dfe0bf9c-en. [1]

Annex B

Annex B. Figures and tables available online

Table B.1. Chapter 1 tables and figures

Figure 1.1		High-scoring holistic creativity responses tend to be highly appropriate
Figure 1.2		Holistic creativity scores and appropriateness scores are strongly correlated
Figure 1.3		High-scoring holistic creativity responses also tend to have high originality and value
Figure 1.4		The most creative answers are more original than they are valuable
Table B1.1	WEB	Share of responses in each appropriate score category, amongst all responses in each holistic creativity score category
Table B1.2	WEB	Correlation between response appropriateness score and holistic creativity score, by task
Table B1.3	WEB	Share of responses in each original score category, amongst all responses in each holistic creativity score category
Table B1.4	WEB	Share of responses in each value score category, amongst all responses in each holistic creativity score category
Table B1.5	WEB	Mean response appropriate and value scores amongst all responses with high originality scores, by holistic creativity score category group
Table B1.6	WEB	Mean response appropriate and originality scores among all responses with high value scores, by holistic score category group
Table B1.7	WEB	Mean response originality score, by holistic score category
Table B1.8	WEB	Mean response value score, by holistic score category
Table B1.9	WEB	Correlation between response originality score and holistic creativity score, by task
Table B1.10	WEB	Correlation between response value score and holistic creativity score, by task
Table B1.11	WEB	Mean idea sum and individual criteria scores for Task 1 - Science Fair Poster
Table B1.12	WEB	Mean response idea sum and individual criteria scores for Task 2 - Illustration Titles
Table B1.13	WEB	Mean idea sum and individual criteria scores for Task 3 - 2983
Table B1.14	WEB	Mean idea sum and individual criteria scores for Task 4 - Save the Bees
Table B1.15	WEB	Mean response idea sum and individual criteria scores for Task 5 - Food Waste
Table B1.16	WEB	Mean response idea sum and individual criteria scores for Task 6 - Save the River
Table B1.17	WEB	Mean response idea sum and individual criteria scores for Task 7 - The Exhibit
Table B1.18	WEB	Variance in holistic creativity scores explained by response appropriateness, originality and value scores

StatLink https://stat.link/5ub4h3

Table B.2. Chapter 2 tables and figures

Figure 2.1		Nearly half of all student responses in 'generate diverse ideas' tasks contained at least one creative idea
Figure 2.2		The share of student responses containing creative ideas varies across country-language groups
Figure 2.3		Only one quarter of responses to 'generate creative ideas' tasks were creative
Figure 2.4		Divergent thinking tasks consistently led to more student responses with creative ideas
Table B2.1	WEB	Share of student responses with none, one, two or three creative ideas, by 'generate diverse ideas' (GDI) tasks
Table B2.2	WEB	Mean response idea sum and individual criteria scores, by task ideation process and by task context
Table B2.3	WEB	Share of student responses with a creative idea, by 'generate creative ideas' (GCI) tasks
Table B2.4	WEB	Share of student responses that include at least one creative idea, by task ideation process
Table B2.5	WEB	Mean idea sum and individual criteria scores of the most creative idea per student response, by 'generate diverse ideas' (GDI) tasks
Table B2.6	WEB	Mean idea sum and individual criteria scores, by response idea position in 'generate diverse ideas' (GDI) tasks
Table B2.7	WEB	Relative share of most creative ideas per student response in each idea position in 'generate diverse ideas' (GDI) tasks

StatLink https://stat.link/7ovijr

Table B.3[1/2]. Chapter 3 tables and figures

Figure 3.1		Few students across country-language groups manage to be creative all-rounders
Figure 3.2		Some country-language groups have more creative all-rounders than others
Figure 3.3		Creative all-rounders are more likely to be girls, be from an advantaged socio-economic background, and score more highly in reading than non-all-rounders
Figure 3.4		Creative all-rounders report stronger attitudes and beliefs associated with creativity
Figure 3.5		Creative all-rounders report higher levels of assertiveness and perseverance, but lower levels of stress resistance
Figure 3.6		Students that report strong attitudes and socio-emotional skills associated with creativity are more likely to score highly across creative tasks
Table B3.1	WEB	Pairwise residual item correlations, by task pairs
Table B3.2	WEB	Share of students that achieved high-scoring responses amongst students that took at least 3 tasks, by number of high-scoring responses
Table B3.3	WEB	Share of girls/boys that achieved high-scoring responses amongst students that took at least 3 tasks, by number of high-scoring responses

StatLink https://stat.link/ks2auh

Table B.3[2/2]. Chapter 3 tables and figures

Table B3.4	WEB	Share of advantaged/disadvantaged students that achieved high-scoring responses amongst students that took at least 3 tasks, by number of high-scoring responses
Table B3.5	WEB	Mean scores in mathematics and reading amongst students that took at least 3 tasks, by number of high-scoring responses
Table B3.6	WEB	Variance in the number of high-scoring responses explained by student gender, socio-economic status, and mathematics and reading scores
Table B3.7	WEB	Mean index scores in select attitudes and traits amongst students that took at least 3 tasks, by number of high-scoring responses
Table B3.8	WEB	Mean index scores in select socio-emotional skills amongst students that took at least 3 tasks, by number of high-scoring responses
Table B3.9	WEB	Strength of the relationship between select attitude/belief indices and mean number of high-scoring responses, before and after accounting for mathematics and reading
Table B3.10	WEB	Strength of the relationship between select socio-emotional skills indices and mean number of high-scoring responses, before and after accounting for mathematics and reading

StatLink https://stat.link/ks2auh

Table B.4. Chapter 4 tables and figures

Figure 4.1		Appropriateness scores contribute most to the average idea score, followed by value then originality scores
Figure 4.2		Ideas in creative expression tasks tend to be more original than valuable, whereas ideas in problem solving tasks tend to be the opposite
Figure 4.3		Most relatively creative responses are highly appropriate, in all country-language groups, though the share varies
Figure 4.4		Relatively creative responses tend to be more highly valuable than original, though not in all country-language groups
Figure 4.5		The relationship between holistic creativity scores and response appropriateness, originality and value scores varies significantly by country-language group
Figure 4.6		Appropriateness, originality and value explain most – but not all – of the variation in holistic creativity scores
Table B4.1	WEB	Share of the mean idea criteria score (appropriateness, originality and value) within the mean idea sum score, by task group
Table B4.2	WEB	Share of responses with a high mean response criteria score (appropriateness, originality and value) amongst relatively creative responses
Table B4.3	WEB	Partial variation in holistic creativity scores explained by mean response score criteria (appropriateness, originality, value)
Table B4.4	WEB	Total variation in response holistic creativity scores explained by mean response score criteria (appropriate, originality and value scores) in 'generate creative ideas' (GCI) tasks only

StatLink https://stat.link/cb9oys

Table B.5[1/2]. Chapter 5 tables and figures

Figure 5.1		Food Waste task
Figure 5.2		Nearly half of all student ideas in Food Waste fall within the same sub-branch of the C-K tree
Figure 5.3		Cross-cultural variability is relatively stable in Food Waste
Figure 5.4		The Exhibit task
Figure 5.5		Over half of all ideas in The Exhibit correspond to two sub-branches of the C-K Tree
Figure 5.6		There is significant cross-cultural variability in the share of ideas across sub-branches in The Exhibit
Figure 5.7		Students found it harder to suggest ideas in different sub-branches in The Exhibit than in Food Waste
Figure 5.8		Share of ideas in sub-branches within the international fixation path for Food Waste and The Exhibit
Figure 5.9		Girls and advantaged students are more likely to suggest ideas within the fixation path than boys and disadvantaged students
Figure 5.10		At least two thirds of ideas within each country-language group corresponds to the international fixation path for the Food Waste task, though the share varies significantly
Figure 5.11		At least two thirds of ideas within each country-language group corresponds to the international fixation path for The Exhibit task, though the share varies significantly
Table B5.1	WEB	Share of students that achieved the maximum response flexibility score on 'generate diverse ideas' (GDI) tasks
Table B5.2	WEB	Share of students by response flexibility score in Task 5 - Food Waste
Table B5.3	WEB	Share of ideas in each sub-branch of the Conceptual Knowledge (C-K) tree for Task 5 - Food Waste (I)
Table B5.4	WEB	Share of ideas in each sub-branch of the Conceptual Knowledge (C-K) tree for Task 5 - Food Waste (II)
Table B5.5	WEB	Modal node and share of ideas (as a percentage of all ideas) within each sub-branch of the Conceptual-Knowledge (C-K) Tree for Task 5 - Food Waste
Table B5.6	WEB	Share of ideas (as a proportion of all ideas) in the 5 most common idea nodes in the Conceptual Knowledge (C-K) Tree for Task 5 - Food Waste
Table B5.7	WEB	Percentage of students with idea flexibility scores 0, 1 and 2 (criteria-based method) in Task 7 - The Exhibit
Table B5.8	WEB	Share of ideas in each sub-branch of the Conceptual Knowledge (C-K) tree for Task 7 - The Exhibit (I)
Table B5.9	WEB	Share of ideas in each sub-branch of the Conceptual Knowledge (C-K) tree for Task 7 - The Exhibit (II)
Table B5.10	WEB	Share of ideas (as a proportion of all ideas) in the 5 most common idea nodes in the Conceptual Knowledge (C-K) Tree for Task 7 - The Exhibit

StatLink https://stat.link/t6wze0

Table B.5[2/2]. Chapter 5 tables and figures

Table B5.11	WEB	Modal node and share of ideas (as a percentage of all ideas) within each sub-branch of the Conceptual-Knowledge (C-K) Tree for Task 7 - The Exhibit
Table B5.12	WEB	Share of students that suggest ideas in 1, 2 or 3 different sub-branches of the Conceptual-Knowledge (C-K) Tree for Task 5 - Food Waste
Table B5.13	WEB	Share of students that suggest ideas in 1, 2 or 3 different sub-branches of the Conceptual-Knowledge (C-K) Tree for Task 7 - The Exhibit
Table B5.14	WEB	International fixation path categories for Task 5 - Food Waste and Task 7 - The Exhibit
Table B5.15	WEB	Relative share of ideas in fixation and expansive paths (international and national path references) amongst boys and girls for Task 5 - Food Waste
Table B5.16	WEB	Relative share of ideas in fixation and expansive paths (international and national path references) amongst boys and girls for Task 7 - The Exhibit
Table B5.17	WEB	Relative share of ideas in fixation and expansive paths (international and national path references) amongst advantaged and disadvantaged students for Task 5 - Food Waste
Table B5.18	WEB	Relative share of ideas in fixation and expansive paths (international and national path references) amongst advantaged and disadvantaged students for Task 7 - The Exhibit
Table B5.19	WEB	Top 5 sub-branches with the highest-scoring mean idea sum scores for Task 5 - Food Waste
Table B5.20	WEB	Top 5 sub-branches with the highest-scoring mean idea sum scores for Task 7 - The Exhibit
Table B5.21	WEB	Mean response idea sum and criteria scores of ideas within the international fixation path for Task 5 - Food Waste, by gender
Table B5.22	WEB	Mean response idea sum and criteria scores of ideas within the international fixation path for Task 7 - The Exhibit, by gender
Table B5.23	WEB	Mean response idea sum and criteria scores of ideas within the international fixation path for Task 5 - Food Waste, by socio-economic status (advantaged/disadvantaged)
Table B5.24	WEB	Mean response idea sum and criteria scores of ideas within the international fixation path for Task 7 - The Exhibit, by socio-economic status (advantaged/disadvantaged)
Table B5.25	WEB	Share of ideas in the international fixation/expansion path for Task 5 - Food Waste, by country-language, region and language groups
Table B5.26	WEB	Share of ideas in the international fixation/expansion path for Task 7 - The Exhibit, by country-language, region and language groups

StatLink https://stat.link/t6wze0

Table B.6. Chapter 6 tables and figures

Figure 6.1		In no country-language group do boys significantly outscore girls on average
Figure 6.2		There are relatively more girls than boys amongst top performers, and relatively many more boys than girls amongst low performers
Figure 6.3		Girls significantly outscore boys in tasks in all domains, on average, and in most country-language groups
Figure 6.4		Girls are more confident than boys at being creative, in general, and telling creative stories, but report far less confidence than boys in creative scientific or invention tasks
Figure 6.5		On average, girls are overrepresented amongst the top 10% of performers in creative thinking tasks
Figure 6.6		On average, boys are overrepresented amongst the lowest 10% of performers in creative thinking tasks
Table B6.1	WEB	Mean standardised holistic creativity score, by gender
Table B6.2	WEB	Variation (standard deviation) in standardised holistic creativity scores of boys and girls
Table B6.3	WEB	Variation (standard deviation) in idea criteria scores
Table B6.4	WEB	Gender differences and gender effect sizes in creative thinking and reading scores
Table B6.5	WEB	Share of boys and girls amongst responses in the top/bottom quartile of standardised holistic creativity scores (all tasks)
Table B6.6	WEB	Mean idea and response criteria scores of boys and girls, across all tasks
Table B6.7	WEB	Mean standardised holistic creativity score of boys and girls, by domain context
Table B6.8	WEB	Mean creative self-efficacy index score and share of students that feel confident in doing select creative activities, by gender
Table B6.9	WEB	Share of boys and girls amongst responses in the top/bottom decile of standardised holistic creativity scores (all tasks)

StatLink https://stat.link/b0lecm

Table B.7 [1/2]. Chapter 7 tables and figures

Figure 7.1		Disadvantaged students struggle the most in coming up with appropriate ideas
Figure 7.2		Disadvantaged students scored comparatively lower than advantaged students in 'generate creative ideas' tasks than in 'generate diverse ideas' tasks
Figure 7.3		Socio-economic status explains a larger proportion of the variation in holistic creativity scores than students' mean response idea sum scores
Figure 7.4		Tasks differ in how much socio-economic status influences student creativity scores
Figure 7.5		Self-efficacy beliefs moderate socio-economic disparities in standardised holistic creativity scores in all country-language groups
Figure 7.6		Students with an immigrant background tend to achieve lower scores than students without an immigrant background
Figure 7.7		Mathematics and reading scores explain most of the differences in scores by immigrant background
Table B7.1	WEB	Correlation between PISA economic, social and cultural status (ESCS) index and standardised holistic creativity scores
Table B7.2	WEB	Correlation between PISA economic, social and cultural status (ESCS) index and mean response idea sum scores
Table B7.3	WEB	Mean standardised holistic creativity scores by student socio-economic profile (advantaged/disadvantaged students)
Table B7.4	WEB	Mean response idea sum and individual criteria scores by student socio-economic profile (advantaged/disadvantaged students) (I)
Table B7.5	WEB	Mean response idea sum and individual criteria scores by student socio-economic profile (advantaged/disadvantaged students) (II)
Table B7.6	WEB	Mean response idea sum and individual criteria scores by student socio-economic profile (advantaged/disadvantaged), by task ideation process
Table B7.7	WEB	Mean standardised holistic creativity scores by student socio-economic profile (advantaged/disadvantaged), by task ideation process
Table B7.8	WEB	Variance in standardised holistic creativity scores explained by student socio-economic status (PISA ESCS index)
Table B7.9	WEB	Variance in mean response idea sum scores explained by socio-economic status (PISA ESCS index)
Table B7.10	WEB	Variance in standardised holistic creativity scores explained by socio-economic status (PISA ESCS index), by task
Table B7.11	WEB	Variance in mean response idea sum scores explained by socio-economic status (PISA ESCS index), by task
Table B7.12	WEB	Average character count of student responses, by task

StatLink https://stat.link/l82vsk

Table B.7 [2/2]. Chapter 7 tables and figures

Table B7.13	WEB	Correlation between standardised holistic creativity scores and creative and mathematics self-efficacy
Table B7.14	WEB	Difference in standardised holistic creativity scores between advantaged and disadvantaged students, before and after accounting for students' self-efficacy beliefs
Table B7.15	WEB	Difference in standardised holistic creativity scores between advantaged and disadvantaged students by task, before and after accounting for students' self-efficacy beliefs
Table B7.16	WEB	Mean standardised holistic creativity scores and mean response idea sum and individual criteria scores, by immigrant background
Table B7.17	WEB	Difference in mean standardised holistic creativity scores and mean response idea sum and individual criteria scores between immigrant (all) and non-immigrant students, before and after accounting for mathematics and reading
Table B7.18	WEB	Mean standardised holistic creativity scores and mean response idea sum and individual criteria scores, by number of languages spoken
Table B7.19	WEB	Difference in mean standardised holistic creativity scores and mean response idea sum and individual criteria scores between students that speak one language and students that speak more than one language, before and after accounting for mathematics and reading

StatLink https://stat.link/l82vsk

Annex C1
Full C-K schema for Food Waste task

This annex is available online at:
https://www.oecd.org/en/publications/seven-questions-about-creativity-and-creative-thinking_0aa52128-en/support-materials.html.

Annex C2
Comprehensive node category list for Food Waste task

This annex is available online at:
https://www.oecd.org/en/publications/seven-questions-about-creativity-and-creative-thinking_0aa52128-en/support-materials.html.

Annex C3
Full C-K schema for The Exhibit task

This annex is available online at:
https://www.oecd.org/en/publications/seven-questions-about-creativity-and-creative-thinking_0aa52128-en/support-materials.html.

Annex C4
Comprehensive node category list for The Exhibit task

This annex is available online at:
https://www.oecd.org/en/publications/seven-questions-about-creativity-and-creative-thinking_0aa52128-en/support-materials.html.

PISA

Seven Questions about Creativity and Creative Thinking

What Do PISA 2022 Data Tell Us?

Student responses on the PISA 2022 creative thinking test are a rich source of data. The CT Rescoring Project applied new scoring methods to PISA 2022 creative thinking data in 14 countries and economies to identify the most creative responses in the sample and shed new light on key questions about creativity and creative thinking. This report examines what makes ideas creative across country-language groups, what cognitive processes lead to more creative ideas, and how successfully students can think creatively in multiple tasks, as well as the characteristics of high-performing students. Policy implications from the key findings are also described.

PRINT ISBN 978-92-64-64690-2
PDF ISBN 978-92-64-94974-4