Routledge
Taylor & Francis Group

Research Papers in Education

ISSN: 0267-1522 (Print) 1470-1146 (Online) Journal homepage: www.tandfonline.com/journals/rred20

The philosophical, methodological and empirical problems in OECD's approach to creativity within PISA 2022

Mark Connolly, Emily Lowthian, D Rhian Barrance & Charlotte Brookfield

To cite this article: Mark Connolly, Emily Lowthian, D Rhian Barrance & Charlotte Brookfield (11 Feb 2026): The philosophical, methodological and empirical problems in OECD's approach to creativity within PISA 2022, Research Papers in Education, DOI: 10.1080/02671522.2025.2610219

To link to this article: https://doi.org/10.1080/02671522.2025.2610219

Published online: 11 Feb 2026.

RESEARCH PAPERS IN EDUCATION
https://doi.org/10.1080/02671522.2025.2610219

Routledge
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

The philosophical, methodological and empirical problems in OECD's approach to creativity within PISA 2022

Mark Connolly[a], Emily Lowthian[b], D Rhian Barrance[a,c] and Charlotte Brookfield[a]

[a]School of Social Sciences, Cardiff University, Cardiff, UK; [b]Education and Childhood Studies, Swansea University, Swansea, UK; [c]School of Social Sciences, Cardiff University Glamorgan, Cardiff, UK

ABSTRACT

Creativity has become a key discourse of the 21st century. Because of this, developing a measure of young people's creative capacity was embedded within PISA 2022. The OECD's approach to creativity is framed within its 'third-phase' of policy-making. This differs from previous economist understandings through the recognition of imagination; a re-engagement with foundational knowledge; and a melding of neo-vocationalist and humanist ideologies. Using mixed methods – qualitative documentary analysis of planning documents and quantitative analysis of data from the PISA Test – this paper argues that the attempt to bring creativity within the purview of PISA measurement is conceptually and methodologically problematic. The paper's documentary analysis illustrates that its 'third phase' framework is conceptually inconsistent and that its creative thinking measurement is theoretically permissive and, consequently, methodologically fragile. The paper demonstrates how this theoretical inconsistency is reflected in the measurement tools and the data from the PISA Test. Consequently, the paper argues that the data has limited explanatory power and utility for policy-makers and educators. We conclude that attempts to develop an international measure using a composite creativity metric is theoretically incoherent and methodologically unrealisable and that the OECD's creative thinking measure could be counterproductive for the promotion and fostering of creativity within education.

ARTICLE HISTORY
Received 20 February 2025
Accepted 13 November 2025

KEYWORDS
Creativity; creative thinking; PISA; OECD; neo-vocationalism; humanism

Introduction

Following the launch of the results of the PISA test in July 2024, it was claimed that the OECD had developed a model for measuring children's creativity. While this 'quest' (Sternberg, Glăveanu, and Kaufman 2024) for a creativity measurement has been ongoing since the 1960s, it has become a focus within strands of creativity research and within the OECD's educational policy in recent years (Connolly 2025). However, as creativity becomes pervasive within policy, what is being referenced is becoming increasingly opaque, militating against attempts to develop robust measurement metrics. This paper's theoretical and philosophical discussions illustrate how this opacity manifests in

CONTACT Mark Connolly connollym4@cardiff.ac.uk School of Social Sciences, Cardiff University, Room 2.26 Glamorgan Building, Cardiff CF10 3WT, UK

the conceptual confusion within both the OECD's 'creative thinking' Framework and PISA 2022's measurement metrics.

The paper begins by considering creativity from philosophical and historical perspectives to illustrate some of these conceptual tensions. Despite the theoretical instability outlined in this first section, there have been attempts to measure creativity which the paper will consider in section two. From this, the paper will trace the role of creativity within different phases of OECD policy-making, where it has evolved from being an economically aligned skill to a 'transformative competence'. [1]The paper will illustrate how this latter conception reframed creativity as essential to both economic and wider human development within its 'third-phase'. Within this 'third phase' of policy-making, the OECD, through its PISA test in 2022, attempted to measure students' creativity across 74 education systems.[2]

In the empirical section, the paper will illustrate how this 'third phase' definition reflects the philosophical tensions discussed in the early section of the paper. This will frame the analysis of the PISA Creativity Test (CT) itself. This analysis focuses on: OECD PISA policy documentation; the development of the 2022 CT; and analysis of the PISA CT results. The paper will end by arguing that while PISA's CT has raised the profile of creativity in education, the OECD's attempt to quantify and rank young people's creativity across global educational systems is theoretically inconsistent, methodologically problematic and empirically fragile.

The history and philosophy of creativity

Research into creativity is riven by methodological and theoretical debates. These reflect more fundamental unresolved philosophical tensions which, this paper will argue, manifest within PISA's CT. This section of the paper will outline these philosophical debates to illuminate its subsequent empirical analysis.

In the study of creativity, we can discern three fundamental tensions. The first of these relates to creativity's association with originality. There are two elements to this: do we need a creative product to be valuable as well as original and what is the relationship between originality and prior knowledge? (Connolly 2025) The second tension emerges from the dichotomy between philosophical positions which argue that creativity is related to a person or process and those which argue that creativity must produce a product (Pirrie 2016). The third tension relates to creativity viewed from a romantic tradition, where it is conceived in opposition to economics (O'Connor 2024) and accounts where it is viewed as the engine of economic progress.

The first of these tensions emerges from creativity's semantic and philosophical essence being derived from its religious and metaphysical associations with 'originality' (Connolly 2025). These residual metaphysical understandings are rooted in theistic debates which grappled with what precedes creation. The e*x-nihilo* (out of nothing) philosophical and theological belief drew on Christian understandings of God's original creation, which was not preceded by anything material: this informed understandings of originality as being aligned with 'newness', but being semantically distinct (Paul et al. 2024).

This association with originality raises an epistemological tension: unlike 'newness' which emerges from a pre-existing frame and is continuous, originality and creativity

subvert or transcend *a prori* frames or precepts and are, fundamentally, discontinuous (Connolly 2025). This understanding not only informed romantic conceptions of creativity, but an anti-intellectualist tradition within progressivist educational ideologies and versions of knowledge-based economy (KBE) thesis articulated within 'second-wave' OECD discourse (Casey 2012). This stress on discontinuity raises one of the tensions within creativity research: if creativity moves beyond established frames and precepts, then how can it be measured? Added to this is a tradition within romantic and aesthetic understandings which retain creativity's association with critiques of economic rationalism and reject both economist appropriations and any attempt to reduce creativity to quantifiable measurement. Creativity's semantic cleaving to originality also raises the question of the possibility of 'nonsense': if originality is the only criterion, then how do we make any value distinctions?

The proposition that a creative person could produce a creative product that is both original and 'nonsense' was first raised by Immanuel Kant (Kant 1790/2013). This has given rise to philosophical debates as to whether originality is enough when judging creativity and the introduction of 'value' to counter the 'nonsense' argument (there are approaches (Boden 2004) which add 'surprise' as a third component, which draws on a tradition of 'epistemological rupture', captured in many genius accounts of creativity in art and science (Connolly 2025)).[3] However, the problems of developing criteria which measure both 'originality' (which is derived from a rejection of *a priori* value criteria) and 'value' (which is built on such criteria) have beset creativity research. If we accept that a value criterion is needed to differentiate creative outputs from novel 'nonsense', then we are left with the question as to how to judge value or, as in PISA 2022, create a universal metric which captures value within such a socially and culturally contingent construct (Connolly 2025).

The second tension arises from understandings of creativity which foreground either the product, process and, in some accounts, the person (Paul and Kaufman 2014). [4]Within process-focussed creativity research, the stress is on psychological capacities such as cognitive flexibility and openness to experience which can foster a creative disposition. They often proceed on the assumption of a link between creative cognitive capacity and its material manifestation. This position has been critiqued from product-oriented understandings (Negus and Pickering 2004) where it is argued that there can never be creativity without a product: Negus and Pickering (2004), for example, conceive creativity as a form of social communication and argue that studies of creativity which do not include a product are only partial.

Debates in relation to value informed distinctions within both the philosophical and psychological traditions which distinguish between individual expressions of creativity and socially meaningful manifestations: the philosophical p/h (psychological/historical) dichotomy; and psychology's Big C/little c categorisation. Drawing on Boden's (2004) work, psychological creativity is conceptualised as novel production at the individual level, while historical creativity represents unprecedented breakthroughs in human achievement. The theoretical genesis of the former -little c- democratised understanding of creativity can be traced to Stein's (1953) seminal work, which initiated a paradigmatic shift from genius-oriented frameworks towards recognition of quotidian creative processes at the individual level. While this served to 'democratise' understandings, without any form of value through social validation, it resulted in a semantic draining and

increased dichotomy between those working within 'little' and 'Big' traditions (Connolly 2025).

Measurement of creativity

These more conceptual considerations underpin the difficulties in developing robust tools for creativity measurement. The first question we must consider is: does the lack of an agreed meaning make creativity measurement an unrealisable, rather than a difficult, ambition? Both Amabile (1996) and Klausen (2010) acknowledge that early research within the field was compromised by conceptual ambiguity. Because of this lack of an agreed meaning, there was a tendency to define creativity through the measure developed to capture the concept. Klausen (2010) identifies this as a form of verificationism, whereby the meaning of a concept is defined by the methods used to gain knowledge *about* that concept. This lack of conceptual clarity can result in inconsistent definitions within tests which can militate against validity of concept; create tautology within questions resulting in circular correlations; and the potential conflation of creativity and general academic ability.

Illustrating these issues, Plucker, Beghetto, and Dow (2004) demonstrated how the majority of papers they analysed did not proffer a definition of creativity to frame their research. Based on this, they argue that much creativity research resembled 'comparing apples, oranges, onions, and asparagus and calling them all fruit' (Plucker, Beghetto, and Dow 2004, 89–90). These problems within creativity measurement were most clearly articulated by one of the leading researchers in the field, Mihaly Csikszentmihalyi, who observed: 'If one turns to the literature of creativity research and asks the simple question: What is being measured? What is creativity? One soon realizes that the entire research enterprise moves on very thin ice' (Csikszentmihalyi 2013 cited in Baer 2016, 11).

These conceptual problems manifest within empirical studies where the distinction between creativity with other concepts 'such as innovation, invention, imagination, talent, giftedness and intelligence' is unclear (Said-Metwaly, Van den Noortgate, and Kyndt 2017, 241). Because of this, much early research has tended to bypass the definitional problem. Where it has been defined, research has often drawn on conceptually contradictory definitions which include: cognitive processes; personal characteristics or dispositions; self-perception; or past learning experiences (Treffinger et al. 2002 cited in Said-Metwaly, Van den Noortgate, and Kyndt 2017). As a consequence, Long et al. (2022) argue that attempts to measure creativity by international organisations such as the OECD may be premature.

Sternberg (Sternberg, Glăveanu and Kaufman 2024) identifies four main strands in the measurement of creativity: cognitive and metacognitive precursor skills; non-cognitive attributes; motivation; environment. As discussed below, PISA's CT attempts to measure all of these: while not directly referencing PISA directly, Glăveanu (ibid 165) warns that these cannot be 'glued together into a creativity score'. Foremost within the field is the first of these: research which has focussed on the generation of psychometric measures. The framing of this research was based on attempts to identify a general concept of creativity through a composite metric which is the equivalent of the *g* of general intelligence (Plucker, Beghetto, and Dow 2004). Guilford's (1975 cited in Said-Metwaly, Van den Noortgate, and Kyndt 2017) process-orientated research aligned

with this approach by linking ideational generation with divergent as opposed to convergent thinking. This dichotomous conceptualisation presents divergent thinking as a cognitive function characterised by the generation of multiple logical alternatives to an open problem; while convergent thought is directed towards a singular, logically determined solution within pre-ordained parameters. Drawing on this, early research focussed on testing for divergent thinking (DTTs) using the criteria of: originality (identified by the uniqueness of responses); fluency (judged by the number of responses); flexibility in the number of categories; and elaboration (the extension of an idea within a category) (Said-Metwaly, Van den Noortgate, and Kyndt 2017, 245). These criteria attempted to capture ideational capacity through respondents' ability to generate multiple and heterogeneous ideas.

There have been ongoing debates in relation to the predictive validity of DTTs. Advocates argue that creativity metrics have stronger predictive validity than conventional intelligence tests in relation to creative achievement: for example, subsequent analyses of Torrance's longitudinal dataset support claims relating to the robustness of DTTs' predictive validity (Runco et al. 2010). However, while DTTs have been used widely within research they have been criticised from a number of perspectives: so much so that Plucker, Beghetto, and Dow (2004) identify them as a 'lighting rod' for critiques of psychometric studies. As Baer argues 'if they are valid measures of creativity, that's wonderful. If they are not, it's a huge problem' (2011, 309).

Even if these correlations are established, Baer (2011) highlights the 'huge problem' within interpretations of Torrance's work which attempt, as in PISA's CT, to devise an overarching creativity measure. Baer (ibid) argues that Torrance actually warned against using the scores within each of his categories to devise a composite creativity metric: to support this, Baer (2011) highlights that Torrance argued that the figural and the verbal had almost no relationship to one another, showing a very weak correlation, leading Baer to conclude that they were not showing different creative modes, but different cognitive abilities.

Consequently, many of the debates in relation to the correlation between the tests and general creative achievement focus on the relationship between DTTs and domain knowledge. The first of these is derived from the epistemological questions raised in the philosophical section of this paper: what is the mediating role of domain-specific expertise in the production of creative thought? Criticisms of DTTs suggest that they are methodologically flawed if they do not account for domain expertise as a critical intervening variable in this relationship: for example, longitudinal research by Weisberg (2006) demonstrates that differences in creative achievement are mostly attributable to prior expertise, while Ericsson's (1999) theoretical framework illustrates that expert practitioners develop domain-specific mental representations which foster creative decision-making.

The initial focus on DTTs within early creativity research has been refined through a recognition of the role of convergent thought within creative thinking processes. An early classic critique by Cropley (2006, 391) revealed 'the seductive promise of effortless creativity' which, he cautioned, can lure one towards a 'pseudo creativity if not adapted to reality'. The process of convergent thinking involves finding an optimal solution to a problem through logic and reasoning. Within creativity research this relates to the phase of thinking where diverse ideas are iteratively refined into workable solutions. This

reframes convergent thought as essential and complementary, rather than antithetical, to creativity (ibid). The importance of convergent thinking in relation to, for example, maths was highlighted by de Vink et al. (2022) who demonstrate that children's mathematical performance is strongly correlated to convergent thinking with limited correlation to capacity for divergent thought alone.

Person-orientated research focussed, instead, on personality attributes associated with creative potential (we will refer to these as dispositional or trait approaches).[5] This research generally focusses on individuals who have been socially sanctioned as creative; identifies shared dispositions or traits; and then attempts to identify these traits, thus indicating a predisposition towards, and a predictor of, future creativity. An extension of this approach is research which attempts to measure attitudes towards creativity itself. Acar and Runco (2015), for example, argue that there are strong correlations between positive attitudes and ideational thinking. These approaches have been linked to the wider concept of 'creative self-efficacy' and personal identity which is seen to correlate with valuing creativity (see Karwowski 2014, for example, in relation to the 'creative mindset').

However, the robustness of self-reporting of dispositional capacities, especially with children, has been questioned: among the limitations are social desirability bias; response styles (either agreeing or disagreeing); lack of insight into one's personality; and cultural differences. Because some of these tests can explore overlapping concepts, their content validity has been questioned: if conceptually weak they can exhibit 'jingle-jangle fallacies', where there is overlap between measurement and concept (Goecke, Weiss, and Barbot 2025) which may result in weak convergent validity between these measures and DTTs (Kagan and Dumas 2025). Another fundamental problem within some studies is that they can exhibit a form of verificationism: they categorise dispositional capacities as a measure of creativity itself, rather than the measure of a disposition which may foster a young person's creativity. Where correlation is explored, the data is disputed: advocates argue that they have construct validity and reliability (Runco, Plucker, and Lim 2001); critics suggest that, while creative individuals frequently exhibit certain characteristic traits, the correlation between personality attributes and measurable creative output remains statistically modest or negligible (Kaufman, Evans, and Baer 2010). In addition, research has shown (Kim and Park 2020) that attitudinal concepts and dispositional awareness are culturally contingent, making establishing a correlation between these and a composite measure of creativity conceptually and empirically challenging.

While such process and person-centred approaches have dominated attempts to measure creativity, they have been criticised from a socio-cultural perspective. These criticisms have argued that such research fails to account for the contextual, and, in particular, cultural factors that inform creative output (Lubart et al. 2019). From this perspective, all creative work is socially and culturally embedded and is thus framed by cultural norms and values: these can inform the skills and dispositions which are promoted; the values which inform personality and self-belief; as well as the expectation of performance (Wong and Niu 2013).

Consequently, within the field of creativity studies, alternative approaches have focussed on the creative product, rather than the person, as the site for measuring creativity. While there is general consensus around the role of novelty and utility in measuring a creative product, socio-cultural research has argued that these concepts are

inherently social and contextual. As a result, several definitions highlight the centrality of social appraisal on the premise that novelty and utility cannot be determined without some form of social evaluation. Thus, for researchers taking this perspective, external judgements of a product would be prioritised over self-report measures of creativity by the person, as we cannot isolate individuals from their social, cultural and historical context (Csikszentmihalyi 2013).

Because of the cultural-contextuality of conceptions of creativity, developing a generic test that is valid and reliable in all contexts is challenging (Glăveanu in Sternberg, Glăveanu, and Kaufman 2024). Some of these empirical challenges include differences in the interpretation of tasks by test-takers and the interpretation and scoring by those marking the test. The latter is of particular importance when the social value of creativity is considered: a product may be ascribed high social value in one context or culture and little in another. Glăveanu (2019, 229), cautioned that, despite attempts to mitigate these cultural factors many creativity tests are still framed within culturally rooted understandings. This raises challenges in developing robust measurement metrics which mitigate cultural biases in relation to concept interpretation and when assessing the value of any creative output or product.

Creativity as skill to creativity as competence

This section of the paper analyses OECD documentation to track the reconceptualisation of creativity within its 'third phase' and, from this, planning within its PISA Test. Within its third-phase, the OECD moved from framing creativity as a discrete skill-based construct, to a more comprehensive competency framework. The OECD's reconceptualisation can be analysed within Xiaomin and Auld's (2020) 'three-phase' typology. Within this typology Phase One (1950–1990) focussed on economic recovery. Phase Two (1990––2010) featured moves towards international comparisons based on neo-liberal forms of governance and neo-vocationalist approaches to education. However, from 2010, there has been a discernible shift away from a neo-vocationalist orientation embedded within a knowledge-based economy (KBE) discourse to a more humanistically orientated 'third phase' (ibid).

Within the former framework, Pappano's (2014) provocation that creativity is no longer 'just for hippies' captures the economic turn which became hegemonic within creativity discourse at the beginning of the millennium. This conceptualisation foregrounded 'innovation', rather than 'imagination', as the key discursive register. The macro-economic framing of this theoretical realignment was late 20th century economic restructuring due to the deindustrialisation of Western economies and the technological automation of procedural labour (Brown, Lauder, and Cheung 2020). The transition from Fordist to post-Fordist production models precipitated a recasting of economic value-creation, where intangible knowledge was foregrounded over tangible skills: as creative-city author Richard Florida (2002, 25) claimed, knowledge workers 'carried the means of production in their heads'. Within this second-phase KBE discourse, influenced by Schumpeterian theories of innovation built on cross-domain knowledge transfer, the skill of creativity was domain-general, knowledge-diverse and economically aligned (Casey 2012). This approach drew upon strong anti-intellectual elements and romantically informed critiques of schools and education systems whose curricula were

organised around a 'collection code' where knowledge has a strong framing within domains and disciplines (Bernstein 1971), which were seen to be anathema to fostering children's innate creative dispositions (this melding of neo-vocational and romantic rejection of strongly framed curricula was encapsulated by the influential Ken Robinson 2011).

Early critics, such as Casey (2012), critiqued the paradoxical relationship with knowledge within KBE discourse. What Casey illustrated was the inherent epistemological tensions in its theorisation of creativity's relationship with domain-specific knowledge: on the one hand, the OECD valorised creativity as a non-domain skill; while, on the other, it promoted what are, essentially, traditional disciplinary and domain-based assessments through its PISA tests. Thus, through its celebration of non-domain skills, the OECD was, in many ways, undermining the validity of its own testing regime. As Grey and Morris (2024) illustrate, the OECD responded in a number of ways. Firstly, it claimed that PISA tests do, in fact, capture these wider skills. To justify this, its second argument was that PISA-successful education systems do not have strongly framed curricula based on a collection code (Bernstein 1971) but, instead follow a more integrated code with weaker disciplinary boundaries. Finally, the OECD signalled that PISA would nuance its tests to capture these wider skills and competencies in future tests (ibid).

These moves were framed, at a more general level, by what has been characterised by Xiaomin and Auld (2020) as the OECD's 'humanitarian turn'.[6] Within this, the OECD has pivoted from a KBE approach to re-engaging with education's role in wider human and social development. This paradigmatic shift manifests in a reconceptualised creativity discourse that seeks to reintegrate both humanist and aesthetic thought, both previously marginalised within neo-vocationalist and economic second-phase frameworks. Within its third phase, the OECD promotes a humanistically oriented conception of creativity as a fundamental competence for contemporary society as well as for future work (Pappano's 2014 hippies were brought back in from the cold). In an attempt to capture these socially and humanistically oriented conceptions of education 'innovative domains' were included within PISA tests from the 2012 cycle, with creativity planned for 2021 and delivered in 2022.

Methods

This paper draws on data from a two-phase mixed-methods study using qualitative and quantitative methods (Cresswell and Plano Clark 2017). The first-phase used documentary analysis of key OECD documents, speeches and reports relating to the development of its PISA CT. The approach to OECD policy documents draws on the concept of 'intertextuality' where all statements are bound up in discursive formations (Fairclough 2010). Drawing on theories of intertextuality and dialogical discursive networks, this research traces the evolution of the discourse of creativity within OECD policy generally but, in its detailed empirical enquiry, within policy documents pertaining to PISA 2022.

The second, sequential phase of the research involved a quantitative secondary data analysis of the publicly available PISA datasets (OECD 2024a). The PISA data was collected between March and December 2022 and features a student questionnaire and tests related to mathematics, reading, science and, of particular interest to this

study, creativity (OECD 2024a). Consequently, the creativity element of the Test had two components: a test for Creative Thinking (CT) and a student questionnaire in relation to attitudes and dispositions towards creativity. To begin, we conducted exploratory analysis of correlations between overall CT scores; overall scores in other components of PISA 2022; and composite scores for the three overarching components within the PISA attitudinal survey. These components were beliefs; attitudes and social/emotional characteristics: for beliefs we averaged each countries' scores for question relating to 'growth mindset and the nature of creativity'; for attitudes we averaged country scores for the questions relating to 'imagination and adventurousness and openness to intellect'; for social-emotional characteristics we averaged scores the questions relating to 'perspective taking; curiosity and persistence'. We then used the PISA approach of ranking countries with the highest score achieving position one, which we then correlated with countries' ranked position for CT.[7]

Our next phase of analysis drew on a sample derived from the 2022 student questionnaire which includes all students who had participated in the PISA tests ($N = 613{,}744$). Of this sample, we linked the creativity dataset ($N = 499{,}843$), and 77% of the participants in the student questionnaire who had creativity data available to be linked. Hence, our sample represents learners who had both student questionnaire data and creativity data available ($N = 499{,}843$).

PISA collected data on creativity through numerous tasks in the domains of written expression; visual expression; scientific problem solving; and social problem-solving situating around generating creative and diverse ideas and the ability to evaluate and improve them (OECD 2024a). There were a battery of 32 items taken by students. Since the tests occurred on a rotated basis, no single participant engaged in all creativity tests, instead working in 30-minute blocks (OECD 2024a).

The metric for creative thinking published by PISA used a combined score of CT capacity. While problems with this, conceptually, are discussed below, methodologically it raises a number of issues, especially with regards to its derivation through the use of plausible values (for a general critique of PISA's use of plausible values see Goldstein 2017). While plausible values can provide a 'measure of performance for each individual on a common continuous scale' (Aparicio, Cordero, and Ortiz 2022, 3342) the complexity of the PISA CT Test makes their use questionable. To create its creative thinking metric PISA generated plausible values using data from all the domains and characteristics within its Test: Barbot and Kaufman (2025) suggest that 70% of the sample did not complete a single creative thinking item. This is especially problematic when the domains from which the creative thinking plausible value was partly derived are being used in correlation analysis.

However, to perform our analysis we had to draw on plausible values to attempt to determine the correlation of student outcomes on differing domains. Therefore, our analysis used the ten plausible values in the domains of creativity, alongside ten each of creativity related to maths, science and reading to produce a correlation matrix. We used Stata 15 (StataCorp 2017) to conduct our analyses, specifically the *repest* package (Avvisati and Keslair 2014) that is able to analyse plausible values whilst accounting for PISA's complex survey design (OECD 2024a). In addition to this, we conducted exploratory factor analysis on Mplus version 8 (Muthén and Muthén 2017) to explore how

a variety of creative tasks perform in relation to capturing underlying constructs of creativity.[8]

Findings

As discussed throughout this paper, the fundamental problematic within the study and measurement of creativity is its polysemy: this manifests in the definitional permissiveness within OECD documentation generally and within the Test itself. This theoretical permissiveness was reflected in many of the planning documents where multiple conceptions circulated. For example, OECD (2019a, 5) claimed that there is a 'consensus amongst educators' that creativity is 'understood as engagement in the thinking processes associated with creative work'; however, where there is no definition of what 'creative work' constitutes this results in a circular logic whereby 'creative work' is defined as the outcome of the creative thinking processes which are defined as preceding creative work.

Consistent with this tautological framing, the Framework argues that the Test will not be a measure of creativity, but a test of the competency of 'creative thinking' which it claims is 'an emerging concept', 'intrinsically linked' to the 'broader concept' of creativity (OECD 2019a, 8). While arguing that 'creative thinking' is a 'tangible' cognitive competence that can be measured, the Framework both endorses and distances itself from 'little-c' understandings: it celebrates the democratic orientation of 'little-c' creativity; but also endorses the socially valuable criterion for creative products which is a feature of 'big-C' and product-derived definitions of creativity. In a dialogue with a debate which we traced to Kant's discussion of the possibility of 'creative nonsense', the Framework argues that 'Creative thinking is thus more than simply coming up with random ideas' (OECD 2019a, 5). To distance itself from dismissals of 'randomness', the OECD adopts the 'novel and valuable' criteria related to creative products.

However, it is in its conflation of process and product that the OECD definition is conceptually problematic. When considering value, the Framework cites the research of Plucker, Beghetto, and Dow (2004, 9) who developed a definition of creativity that is both product and socially orientated, where novelty and value are socially determined:

> the interaction among aptitude, process and environment by which an individual or group produces a perceptible product that is both novel and useful as defined within a social context.

What the precise definition forwarded by Plucker, Beghetto, and Dow (2004) attempted to address was the lack of semantic and conceptual rigour within strands of creativity research. What is key in Plucker's, Beghetto, and Dow (2004) definition is that it is through the production of a perceptible product whose value has a social imprimatur, that we can then consider the individual, social and cultural contexts which have fostered the creative person who developed this creative outcome. However, appropriating this definition, the OECD articulated a definition of creativity as:

> competence to engage productively in the generation, evaluation and improvement of ideas, that can result in original and effective solutions, advances in knowledge and impactful expressions of imagination. (OECD 2019a, 8).

Unlike Plucker's, Beghetto, and Dow (2004) definition (where the process is justified by the product), the focus here is on the competence of 'creative thinking' itself which *can* result in a product. The products which may emerge reflect the OECD's third-phase discourse which includes solutions, knowledge advancement and the aesthetic discourse of imagination (OECD 2019a) (in the Test itself these expressions of imagination were through a computer-based drawing tool with limited utilities or affordances for imaginative expression). Although part of the measurement was through a product, the OECD definition attributes value to the *process*, where value can be discerned as 'productive engagement' in the process of idea generation and improvement itself. Within this definition and subsequent Test, productivity is mainly measured through ideational fluency. This represents a fundamental issue within the OECD definition, where process and product are 'glued together' and any correlation and potential causality jettisoned within the concept of 'facets'. These facets are: 'generate diverse ideas' (weighted at 40%); 'evaluate and improve ideas' (30%); and 'generate creative ideas' (30%) (OECD 2022, 22- see Figure 1 below). While one may or may not support DTTs, the first two of the facets are rooted in that tradition and map onto the first clause of the OECD definition.

The third facet, 'generate creative ideas' aligns to the second clause of the OECD's definition. Despite its ideational framing, this facet is theoretically rooted in product-orientated understandings of creativity: 'the creativity of students' products provides indicators of their capacity to think creatively, particularly in tasks where much of the creative thinking process is 'invisible'' (OECD 2019a, 17). To rank the creativity of students' products, the OECD uses the criteria of 'useful', as well as 'novel', to counter the Kantian nonsense argument (ibid). Arguing that 'total uniqueness' and societal validation are both inappropriate within the Test, the OECD introduces a metric for novelty defined by 'originality'. This understanding of 'originality' is divorced from its romantic associations and, instead, defined by 'statistical infrequency' based on 'deviance from patterns observed within the population at hand' (ibid). Value is defined through the indistinct concept of 'appropriateness', a criteria which 'means that the response must comply with the basic requirements of the task, respect the task constraints (if present), and reflect a minimum level of usefulness in the response' (ibid) (though the actual guidance given to the students when taking the test only stresses that their answers are original).

This attempt to incorporate theoretically and methodologically diverse perspectives features throughout OECD Framework and planning documentation. Within the Framework's expansive account, the already overextended definition is stretched further to claim that creativity can improve 'metacognitive capacities, inter- and intra-personal and problem-solving skills, as well as promoting identity development, academic achievement, future career success and social engagement' (OECD 2019a, 5). This discourse, emanating from the Strategic Advisory Board, cites the inflated creativity rhetoric within policy emanating from the UK government at the turn of the millennium, where creativity was posited as a nostrum for the country's economic and social maladies (see Connolly 2013, 2025).

A key element within this expanded definition was the OECD's re-engagement with the discourse of imagination within its 'third-phase'. This reconciliation was evident in Andreas Schleicher's address (Schleicher 2019), when introducing the *Learning Compass 2030*. Schleicher invokes a romantic understanding of creativity

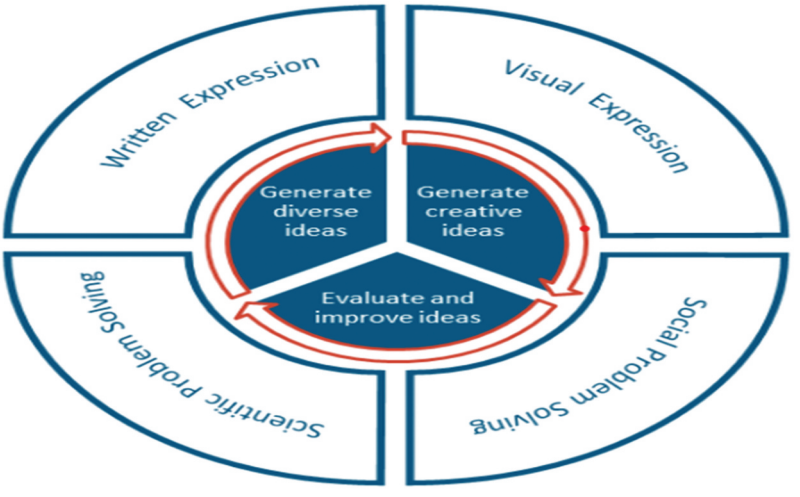

Figure 1. PISA CT Framework. The outer sections represent creative domains; the inner sections are facets of creative thinking (OECD 2019a, 23).

and introduces the word 'aesthetics' into the OECD's formerly economist discourse. This address exemplifies the OECD's third-phase shift through its emphasis on human's creative capacities forged on a techno-dystopian narrative which positions human agency as an affective bulwark against AI-driven social fragmentation. However, the epistemological essence of this aesthetic discourse is that creative thought is the antithesis of rational thinking: as Connolly (2025) outlined, a key discourse within romantically informed accounts of creativity in education from the end of the 19th century is that 'schools kill children's innate creativity'. The rejection of integrated curricula which emerges from this manifests within the OECD's planning documentation. In an early mapping document (OECD 2019a), the OECD draws on the 'schools kill creativity' trope by invoking the portmanteau 'creaticide' in relation to education systems which are seen to suppress, rather than foster children's innate creativity.

However, while invoking this trope, the OECD paradoxically validates knowledge, domains and disciplines as the roots of creativity: rather than creativity being seen as a non-domain skill, it is, instead, rooted in subject disciplines within curricula: 'Schools naturally have an important role to play in developing children's domain readiness (knowledge and experience) in a range of subject areas in which students can express their creative thinking' (ibid, 13). To paraphrase the title of Michael Young's seminal work in this area, the OECD are attempting to 'bring knowledge back in' (Young 2008). This is explicit within the Framework where Gardner's conception of 'foundational' knowledge is cited, as well as the endorsement of the critical realist and 'powerful knowledge' epistemologies, particularly the work of Muller and Young themselves (OECD 2019a). The endorsement of the 'powerful knowledge' perspective is illustrative of the OECD's ambivalent relationship with knowledge and domains, as it represents a critique of the epistemological foundations upon which 'integrated' curricula (Bernstein 1971), which the OECD promoted, are built.

These conceptual contradictions in relation to knowledge can be illustrated in how the OECD frame the concept of a domain itself. Quoting Baer, the Framework defines a domain as 'any specific area of knowledge, such as art literature, history, or astronomy' or 'the set of representations that underlie and support thinking in a specific area of knowledge' (Baer 2011, 27). As Baer outlines, the definition of a domain has an epistemological framing along Bernstein's (1971) 'collection code': it is an area of knowledge or a set of representations which are shaped by particular epistemological foundations. Using this epistemological definition the Framework recognises the domain-specificity of creative thinking which it illustrates through differences between writing a poem and generating scientific hypotheses (OECD 2019a, 23). However, throughout the Framework and the wider CT test, there is ambiguity and inconsistency in both defining this fundamental concept and in relation to the domain-specificity of creative thinking. While using Baer's definition, the Framework also cites accounts of domains which adopt a much broader conceptualisation which do not use the epistemological framing but a more general understanding of domains: everyday; scholarly; performance; scientific; artistic. This general understanding is then related to Amabile (1996) who uses the categorisation of verbal; artistic and problem-solving (cited in OECD 2019a). Domains are then categorised as two 'thematic content areas' - 'creative expression' and 'knowledge creation and problem-solving'- which are subdivided into the 'domains' of 'written expression', 'visual expression', 'social problem-solving' and 'scientific problem-solving' (ibid, 23- see Table 1).

'Problem-solving' is defined within the Framework as both: a. non-domain skill; and b. a discrete domain. While in other parts of the document the Framework suggests its interpretation is a. 'creative problem-solving' i.e. 'finding creative solutions to a variety of problems across domains' (11); within the Test, creative problem-solving is represented as b: 'social problem solving'; and 'scientific problem-solving'.

The fundamental question then arises as to how and why the scores within these domains were fitted into one scale to create a general metric for creative thinking: what is it that this metric is actually meant to represent and what contribution does

Table 1. Thematic content areas and domains within PISA ACT framework (taken from OECD 2019a).

Thematic content area	Domain	Percentage of total items
Creative Expression	Written expression	25%
	Visual expression	25%
Knowledge creation and problem solving	Social problem solving	25%
	Scientific problem solving	25%

Table 2. Exploratory factor analysis of tasks in PISA CT Test.

Task	Loadings
Science fair poster	0.42
Space comic	0.44
Clean ocean	0.58
Experiment kit	0.63
Save the river	0.57
Illustrative titles	0.37
Save the bees	0.65

the ranking of countries based on this metric make to our understandings and promotion of creativity within school? While this may be methodologically and conceptually problematic, it is strategically and politically important to the OECD: the ranking imperative based on a single metric is fundamental to the OECD's leveraging of political change through mediatised post-PISA failure narratives (Grey and Morris 2018).

To complement this multi-faceted and complex CT test, the OECD also formulated a dispositional survey to capture students' dispositions and attitudes towards creative activity. As with the Framework and the Test, this attitudinal survey lacks conceptual precision. Goecke, Weiss, and Barbot (2025) argue that while all creativity measures involve conceptual challenges, the international and cross-cultural nature of PISA makes issues around content validity even more pronounced and can result in conceptual confusion which may compromise the survey's validity (this is illustrated in detail below).[9]

The empirical problems emerging from this maze of complex measurement was recognised and pre-empted by the OECD itself when it states: 'scoring challenges are greater for this assessment than for any other PISA domains' (OECD 2019a, 28). The next section of the paper will illustrate how these scoring challenges manifested in the data generated from this CT Test.

These challenges manifested immediately in the reporting of the CT result in 2024. When introducing the results, Andreas Schleicher articulates a version of the 'schools kill creativity' trope (Schleicher 2024). Schleicher justifies this claim through data which suggests that children become less creative as they age: he states that, on average, there is a substantial drop in children's creativity between the ages of 10 and 15 years. This results in him asking: how would we react if '15-year-olds do worse that 10-year-olds in mathematics' (Schleicher 2024).

However, this introduction illustrates the theoretical and methodological issues in defining and measuring creativity discussed earlier in this paper. Firstly, this data is *not* from the CT which PISA is presenting, but from a previous survey based on young people's creative *self-perception* where creativity was seen as a social and emotional skill (SES) (OECD 2024d). The 'schools kill creativity' trope was first forwarded within this SES survey when it argued:

> Ten-year-olds often exhibit a natural curiosity … as they mature, they are compelled to conform to adult modes of thinking, which can result in diminished innate creativity. (OECD 2024d, 25).

Certainly, these changes in children's self-perception during adolescence are interesting, as are the differences in perception between countries and cultures. However, Schleicher (2024) introduces a series of problematic assumptions based on this: 1. that this data relates to creativity as a cognitive skill, rather than *perception* of a social and emotional capacity; 2. that this is evidence of a decline in a cognitive skill; 3. that schools are the cause of this decline; 4. that school pedagogy enforces a compliance which destroys children's innate creativity.

> … school students should always learn something. But, you know, maybe not … we see a steep drop in creativity … when we send children to school we try to make them 'compliant' with our ways of thinking. (Schleicher 2024).

Categorising creativity as a social and emotional skill (SES) raises its own issues, especially since both its definition and its behavioural manifestation differ within the two OECD surveys. Within the SES survey, creativity is defined closely and almost exclusively to art with the behavioural manifestation being defined as having 'original insights' and creating 'valued art works'; while the inverse is defined as acting 'conventionally and not interested in arts' (given this definition it is unsurprising that the survey discovers a correlation between creativity and studying arts (OECD 2024d,12)).

So, the OECD begins its presentation by drawing on a survey of children's self-perception of their creativity, defined as a social and emotional skill and measured through a self-perception questionnaire. This draws on a definition and manifestation of creativity most closely related to the arts; which it then argues is evidence of an inverse relationship between creativity and schooling; which it then uses to frame a study which views creativity as a cognitive competency-creative thinking; which uses an alternative definition and measurement of creativity/creative thinking most closely related to problem solving. The 'thin ice' that Csikszentmihalyi (2013) claimed measures of creativity are built upon would seem to become even more fragile within the OECD's analysis.

The fundamental epistemological tension between creativity and domain and disciplinary knowledge in schools further undermines the robustness of the PISA CT and questions over what it 'is actually measuring' (Csikszentmihalyi 2013). Using a variety of metrics, domain knowledge scores within PISA tests and CT scores are strongly correlated. This can be illustrated at a system level through an analysis of education systems' ranking in both the PISA maths, science and literacy tests and their ranking in the PISA creative thinking test. What we can see from this is that 90% of top twenty countries in maths, literacy and science are also top twenty countries in creative thinking. Where movement did exist, it was at the margins: two countries moved in to positions 19 and 20 while one country- Hungary- moved four places from 19–23 (the non-marginal exit was Slovenia which moved from position 11–25). There was no movement in positions 1 and 2 with Singapore holding position 1 and Korea position 2 (Japan who held position 2 in maths, science and literacy did not partake in the creativity element of PISA); while 85% of the countries in the top 20 in maths, reading and science were within 5 places when they were ranked for creativity (OECD 2024c).

This relationship was further reinforced within Schleicher's (2024) presentation, where the OECD correlated systems' scores in Maths with their CT score. The first question when considering this data might be: why would you perform this correlation? Conceptually, this framing is problematic: since there is no test of 'mathematical creative thinking', why correlate a composite CT score with domain of maths? Is maths being used as a proxy measure for 'general domain-based knowledge'? [10] If we did accept this (which we do not), having established a strong correlation between maths scores and creative thinking scores, the conclusion would be that creativity is firmly rooted in the domain of maths and emerges through knowledge immersion. This reinforces the janus-faced relationship the OECD has with domain knowledge and schools outlined above: how can there be *both* an inverse correlation between a child's age and their creative capacity (that they get *less* creative as they get older because of schooling as Schleicher argues); and a strong positive correlation between a child's ability in mathematics (which improves as they get older and is learnt in school) and their creativity?

Following our own analysis of the PISA's creativity index through using plausible values of creativity generally, and in the domains of maths, reading and science, we found that there were strong positive relationships between students' PISA creativity scores and all three domains of reading, science and maths. We explored the creativity measure's relationship with the general subject domains (outside of the creativity domain dataset) and found remaining strong relationships in reading (0.72), maths (0.71) and science (0.71). Our findings align with those which PISA reported (OECD 2024c), whereby students who score highly in the creativity tasks had high correlation with both the general domains of maths, reading and science. Our data shows a slight increase (0.04 - 0.05 unit difference) in value compared to PISA data, which is possibly due to slight differences in methodology (as we only can access the 2015 methodology or software packages). While, historically, PISA has resulted in high correlations between its domains of maths, science and reading (resulting in some of the critiques discussed above) these strong correlations raise a number of questions in relation to the method of PISA and in relation to how these results were interpreted and reported in its subsequent analysis.

To nuance this strong correlation, the OECD drew on the relationship between quintile positions in maths and creativity, focussing on the top quintile. Recognising that over 50% of the top quintile were the same students and 58% of the bottom were also the same, the OECD argued that maths and creative thinking go 'hand in hand' (OECD 2024b, 86). Rather than focussing on the strong correlation, however, the OECD argued that these correlations proved that there was something happening *other* than domain knowledge in maths. This might be justifiable if they were correlating between maths and 'creative maths' for example or if, more fundamentally, there was a consistent epistemological framing of domains (this becomes even more problematic through the use of plausible values where the creative thinking metric was partly derived from scores in the domains of science, reading and maths (Barbot and Kaufman 2025). Instead, the OECD analysis concentrated on the movement of students within the third and fourth quintile to argue that academic excellence was *not* a 'pre-requisite' for creativity.

We would, however, question this interpretation from a number of perspectives. Firstly, as discussed above, without a consistent epistemological framing and domain-based assessment which related to creativity in maths, establishing a correlation is problematic (as Plucker, Beghetto, and Dow 2004 cautioned we cannot compare asparagus and apples and call them all fruit). Secondly, if we accepted that these correlations were meaningful, then given the strong correlations, the consistency in systems' positions and even the relative stability in quintile positions, the conclusion would seem to be that foundational knowledge *was* a 'pre-requisite' for creativity. To argue that foundational knowledge in maths, science and literacy is fundamental to creative thought would then lead to the promotion of a curriculum based on a collection code (Bernstein 1971), which the OECD rejects.

Instead, could the marginal differences be attributable to students who score highly in maths not scoring highly in the expressive component of the creativity test? The OECD's own planning documentation would seem to offer this explanation where rather than advocating for a composite measure of creativity, it argued that creativity in the epistemologically framed domains of maths and art drew on different sets of resources: 'In particular, research has found that creativity in the arts and creativity in maths/science

draw upon a distinctly different set of internal resources (e.g. knowledge, skills, and attributes)' (OECD 2022, 11).

However, because of the limitations of the dataset (particularly the use of plausible values- see above) we were unable to perform this more detailed analysis of scores within the visual expression facet and overall scores in maths. In an attempt to assess whether creativity tasks were capturing separate constructs or domains, as PISA intend, we conducted an exploratory factor analysis. This allowed us to understand if student performance on certain tasks align stronger than others, or if all items were capturing student performance in a singular manner. We used the creativity tasks where there was sufficient sample available ($N = 98,677$), as all students did not do all tasks. The tasks we included differed in their test aims which included various domains and facets and included as follows:

- Science fair poster
- Space comic
- Clean ocean
- Experiment kit
- Save the river
- Illustrative titles
- Save the bees

Using a cut-off of 0.40, only the illustration titles (0.37) did not load well on to a single factor, although it was only marginally missing the cut-off and other researchers may well include loadings ≥0.30. The one-factor model was the best fitting model (RMSEA = 0.01, CFI = 1.00, TLI = 1.00, SRMR = 0.02) with an eigenvalue above 1.00. While we included a variety of tasks across the domains, this analysis suggests that there is an overarching construct being captured across all tasks included here. Given the tasks are rooted in different domains and measure different facets of creativity, it would not be expected that they would all load in a single construct. The one task which did not covary was *Illustrative Titles* rooted within the domain of 'visual expression' (see Table 2 below). While we can only offer tentative conclusions from this, it does suggest that differences in scores may, at least partly, be due to the skills in this domain differing from other domains. Whilst significant work went into PISA's piloting of these items (OECD 2024a), further research must be conducted on the sample to explore if the responses truly capture the constructs intended. Moreover, it is unclear if these constructs hold over country and cultural contexts, and measurement invariance of creativity items should be explored across countries (Boz 2025), particularly as many focus on DTTs (Guo et al. 2024).

Our findings not only indicate issues with correlations between domain knowledge and the PISA CT score but between this metric and data from its dispositional survey. As the previous discussion illustrated, there have been ongoing debates in relation to studies which have focussed on measuring creativity through such traits or dispositions. Many of these issues emerged within the dispositional dimension of the PISA 2022 creativity survey. While the results from this section of the survey may be informative in relation to illuminating cultural understandings of concepts such as persistence, curiosity or attitudes towards creative thinking, their utility in illuminating how education can foster

creative thinking is questionable. This can be illustrated most clearly by the negative correlation ($r = -0.25$) between CT and in-school creative activities (OECD 2024b). The conclusion from this would be that to promote CT in schools we should reduce creative activity.

We can illustrate this further by considering three of these traits: persistence; curiosity; and 'attitudes towards creative thinking'. If we consider the first of these – persistence – 60% of the countries with the highest score are from Latin or South America. What this might tell us is that persistence is culturally valued in these countries and is emphasised within its education system and that students thus *perceive* themselves to be persistent (OECD 2024c). We can identify a similar clustering in relation to curiosity where, again, 60% of the most 'curious' students within 'systems/economies' are from South or Latin American countries.

However, the most problematic attitudinal data relates to students' attitudes towards creative thinking itself. Interestingly, within this question, the actual definition of creative thinking as 'openness to intellect, art and experience' differs from the definition outlined in the Framework document itself and the SES survey cited above. The countries which have the most positive attitude towards creative thinking are Albania, Thailand and the United Arab Emirates. If we cross-tabulate these attitudinal, dispositional and trait scores to these countries' creativity scores based on the PISA creative thinking index, we can actually detect an inverse, rather than a positive correlation. The most striking example is Albania where students are said to have the most positive attitude towards creativity (mean index 0.40 'openness to intellect'; 0.50 'openness to experience') of all countries; while, at the same time, the lowest creative thinking score (score 13; position 37). Of the ten countries where students were deemed to have the most positive attitude to creative thinking, 6 were in the lowest quartile; four in the third quartile and 0 were in the second or first quartiles for the overall score of creative thinking (OECD 2024a).

We conducted further analysis to probe this inverse correlation between countries' position in CT and a composite measure of position metrics relating to attitudes towards creativity as illustrated in Figure 2.

Figure 2 demonstrates the correlation between countries' creative thinking positions (x-axis) and students' attitude towards creativity (y-axis) where 1 indicates the highest score. This illustrates the negative correlation between countries' overall position for attitudes towards creativity and its composite CT score. As the bottom left quadrant indicates only one country in the top 20 for creativity was also in the top 20 for attitude (position 19); while seven countries in the top 20 for attitude towards creativity were ranked lowest (position 50–60) for CT.

We can identify similar patterns for persistence and curiosity. Of the ten countries where students are deemed most persistent only one is in the first quartile for CT; 2 are in the second; 3 are in the third quartile; while the majority, 4, are in the fourth and lowest quartile (OECD 2024c). To investigate this further we created a composite score for countries' positions relating to curiosity and persistence which we correlated with their position for CT.

Figure 3 illustrates this correlation between countries' CT scores (x-axis) and students' curiosity and persistence scores (y-axis) using the same ranking where 1 indicates the highest score. While not as distinct as the negative correlation between attitudes and creativity, there is no indication of a positive correlation and some of

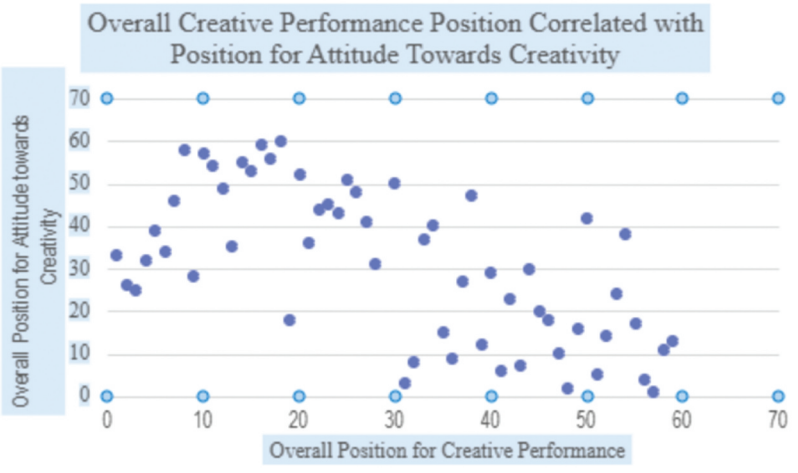

Figure 2. Overall position for creative thinking correlated with position for attitude towards creativity.

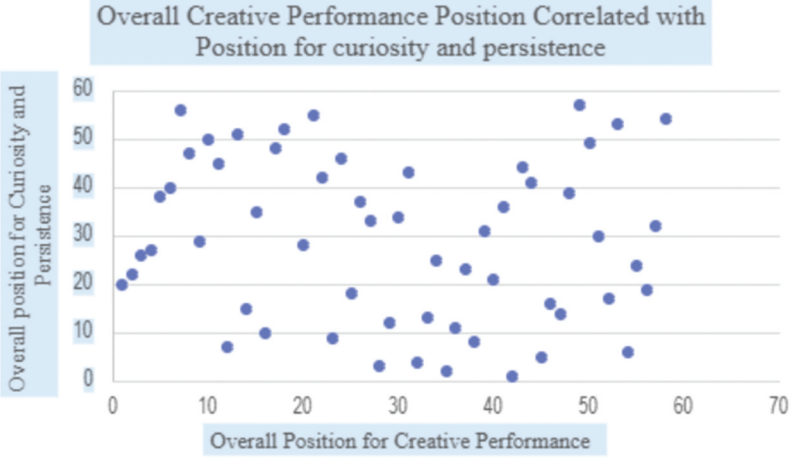

Figure 3. Overall position for creative thinking correlated with position for curiosity and persistence.

a weak negative correlation. As indicated in the bottom left quadrant no country in the top 10 for Creative Thinking ranked in the top 10 for curiosity and persistence, while a majority countries in the top 50% for creativity were in the bottom 50% for curiosity and persistence.

When analysing this component, the OECD claims that 'The four indices of attitudes towards creative thinking are positively correlated with one another' (OECD 2024b, 163). This reporting reflects the conceptual ambiguity, theoretical imprecision and methodological fragility which we have identified in other elements of the PISA Creativity Test. The indices used to measure attitudes within the Test are: Imagination and Adventurousness; Openness to Intellect; Openness to Art and Experience; and Creative Self-efficacy. An examination of the measures used within these indices suggests that many of the metrics were measuring

similar concepts: a classic example of the 'jingle' in the jingle-jangle fallacy discussed above.

This can be illustrated by examining the OECD's 'Openness to Intellect' index. Originally conceived to identify creative thinking aligned with scientific enquiry this construct was operationalised in PISA 2022 through questions which seem unrelated to this conceptual framing, such as: 'doing something creative satisfies me'; 'I am very creative'; 'I like creative stories'. The OECD then reports correlations between this index and other measures: Creative Self-efficacy ($r = 0.54$), Openness to Art and Experience ($r = 0.49$), and Imagination and Adventurousness ($r = 0.47$). The strongest correlation, between Openness to Intellect and Creative Self-efficacy ($r = 0.54$), illustrates these methodological issues. The self-efficacy questions include: 'I enjoy coming up with creative ideas'; 'I enjoy being creative'; 'I enjoy telling creative stories'; ' 'I enjoy making creative drawings.' These items are virtually identical to those measuring Openness to Intellect, making strong correlations inevitable and empirically problematic. Similarly, within the dispositional element of the survey the OECD reports correlations between stress resistance and emotional control ($r = 0.51$) and between persistence and curiosity ($r = 0.37$): this is unsurprising given that these represent closely related concepts measured through similar questions (even if these were robust, finding a correlation between these does little to inform our understanding of creativity). The Report then uses these correlations to make sweeping claims such as 'In a nutshell, the most open-minded students are the most imaginative (and vice versa, with all four attitudes' OECD 2024a, 163) which, as illustrated above, is empirically and conceptually problematic.

Conclusion

To begin we must recognise the positive elements within the OECD's approach to creativity within its third phase and its attempt to develop a measure within its PISA 2022 Test. The OECD's foregrounding of creativity as a distinctly human characteristic that is beyond the capacity of machine intelligence was both prescient and apt. Both the humanist turn within its policy and the abandonment of purely economic justifications of creativity is to be welcomed. While the inclusion of creativity within its PISA test raises issues, it does draw attention to creativity research and the importance of fostering children's creative potential within schools.

While these aspirations are laudable, they may also have sown the seeds which undermined the OECD's attempts to measure creativity: as the OECD itself recognises, measuring creativity is difficult and its test complex. This methodological complexity was, in part, the consequence of the adoption of an inclusive account and the attempted rapprochement between differing philosophical perspectives, theoretical traditions and methodological approaches. However, this did not resolve these fundamental debates; instead, it lay the ground for the epistemological tensions, theoretical inconsistencies and empirical weaknesses which manifested within the CT test.

This conceptual latitude resulted from the OECD drawing on inconsistent definitions of creativity within its own studies, resulting in what the paper described as 'definitional permissiveness'. One of the criticisms of creativity research has been its definitional latitude (Plucker, Beghetto, and Dow 2004). However, the OECD drew upon competing definitions, not only within its wider documentation and other research, but within the planning for and the reporting of the data from the Test itself. This theoretical

promiscuity and conceptual ambiguity was a feature of both the Test and the supplementary questionnaire which used inconsistent definitions of creativity; forms of verificationism within its metrics; and overlapping questions and measurements to capture distinct concepts. This raises questions in relation to the robustness and utility of the data as well as the OECD's interpretation of the results.

One fundamental unresolved epistemological tensions within this Test and within OECD educational policy generally relates to knowledge and domains. Despite the rhetorical rapprochement with disciplinary knowledge within the OECD's third phase, the residual anti-intellectual strand within second-phase discourse informed the development of the Framework and the reporting of the results: as with the OECD's general policy, it both embraced and distanced itself from domain knowledge. This was most clearly illustrated in relation to its rearticulation of the 'schools kill creativity' trope which formed a key discourse within the promotion of creativity as a non-domain skill within the OECD's second phase. This resulted in the OECD making the theoretically and empirically inconsistent claim that, on the one hand, schools kill creativity and that children become less creative as they get older while, on the other, arguing that domain knowledge (taught in schools) is fundamental to creativity. Schools cannot both kill and foster creativity; and children cannot become both less and more creative by attending school.

While the PISA 2022 framing document cited researchers who critiqued the use of domain-general and composite creativity measures, its test of 'creative thinking' developed a non-domain composite metric through which it ranked countries' educational systems: such non-domain conceptualisations is one of 'garden paths' which Sternberg (2022) warns creativity researchers from wandering down. While there is certainly value in gathering data on culturally mediated understandings of creativity across the globe, we question the use of measuring and ranking countries through the generation of a composite and reductionist measure of creativity which subsumes the rich social and cultural contextuality of creative work into a standardised universal metric. This questioning resonates with ongoing debates in relation to PISA, not only in relation to its methods but also its policy impacts. One area of debate is the purpose of ranking countries within PISA's metrics which critics argue generates media attention and, through this, leverages policy change (Grey and Morris 2018). While a composite measure of creative thinking may have realised this goal it did so at the expense of facilitating nuanced understandings of children's creativity.

As with the concept of creativity itself, the OECD's definition of the concept of a 'domain' militated against developing a robust and conceptually coherent framework: the paper illustrated that the definition of 'domain' was theoretically ambiguous and diverged from the epistemological framing within its own Framework and other PISA tests. The OECD correlated its CT score with its tests in maths to argue that creative thinking was a distinct concept. However, countries' positions in other PISA tests were consistent and strongly correlated with positions in its CTT. We argue that there are two conclusions that can be drawn from this. Firstly, if we accept that these tests were measuring distinct concepts then the OECD should be highlighting the strong relationship between disciplinary and domain knowledge and creative thinking skills. The second explanation is that both tests were measuring the similar cognitive capacities which our exploratory analysis suggests may be the case. This is an area that needs further

detailed analysis before any robust conclusions can be made. The use of measures where correlation is not been established is another of the paths that Sternberg (2022) warns against treading. This issue of correlation was most apparent in relation to the dispositional survey which complemented the CTT where there was a negligible and, in some cases, an inverse correlation with the OECD's composite creative thinking metric. Most striking was the negative correlation between participation in creative activities and PISA's CT score. If we were to accept this, we would promote creative thinking by reducing creative activity in school.

The OECD argued that the main objective of the Test was to 'to provide internationally comparable data on students' creative thinking competence that have clear implications for education policies and pedagogies' (2019b, 18). This paper has argued that the data itself should be treated with caution; that the concept of creative thinking needs further refinement; and that there it has limited utility in relation to developing policy or pedagogy which will foster and promote creativity in schools. While research into how creativity is understood and manifests in countries and cultures is to be welcomed, the OECD's attempt to generate a composite international measure through which countries are ranked may not only be unachievable but may also be counterproductive in promoting creativity within global education.

Notes

1. The alignment between innovation and creative thinking was, however, retained in other OECD publications during this period (Vincent-Lancrin 2019).
2. 2.81 countries participated in the PISA test overall. Of these 64 countries took the Creativity Test component while 74 took the dispositional questionnaire in relation to creative aptitudes and dispositions (OECD 2024b).
3. It seems as if the OECD rejected this as its data analysis of the PISA Test it discounted any answer that was generated in under 15 seconds on the grounds that 'creativity needs thought' (OECD 2024c).
4. The three or four 'p' (where press is added) framework has dominated creativity research. There have been attempts using a sociocultural perspective to reframe this as a five A framework where more attention is given to the interrelationship between dimensions (Glăveanu 2013).
5. The nomenclature has been debated with some nuance in how the words trait and disposition are used. Lucas, Claxton, and Spencer (2013) acknowledge that Guilford uses 'trait' but they use the word disposition. In this paper we try to be consistent with the nomenclature used by researchers themselves.
6. We have developed this recognition of a humanitarian element in OECD policy to a broader understanding of 'humanism', where human values are seen as a counterbalance to threats posed by artificial intelligence.
7. We only included countries with data relating to all of these measures which meant that 13 countries which participated in PISA 2022 Creativity Test were not included in this analysis. Unfortunately, creative thinking data for countries which we were particularly interested in – UK; Ireland and the US- was unavailable. Barbot and Kaufman (2025) suggest that this was not only because of policy priorities but concerns about how the measurement would work.
8. We note that throughout this analysis, the usability of the datasets provided by PISA are considerably complex and are not user-friendly for quantitative social scientists. The documentation provided is often not clearly communicated and the survey design technical guidance on data availability of the creative items is not easily accessible. We argue that

significant improvements should be had by PISA to encourage quantitative social scientists to use this potentially valuable dataset.

9. For a detailed discussion of how this lack of precision compromises content validity see Goecke, Weiss, and Barbot (2025).
10. Although strongly correlated, maths had a weaker correlation than reading. One may interpret this as the OECD wanting to show weaker correlations to bolster its argument that the CT test was measuring different cognitive capacities. Within its technical documentation the OECD (2024d) argued that the use of maths was based on the availability of data.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Dr. Mark Connolly is a Reader in Education in the School of Social Sciences and interim Director of the WISERD Education Multi-Cohort Study, a longitudinal cohort study of secondary school children in Wales. His theoretical interests include education, cultural policy and creativity. His other research focuses on range of educational projects relating to outdoor and Forest School learning; routes into teaching; teacher education; professional leadership; children's rights; children's play. Mark has been the Welsh representative on the council of the British Education Research Association and is a former editor of the British Educational Research Journal.

Dr. Emily Lowthian is a Lecturer at Swansea University in the Department of Education and Childhood Studies, School of Social Sciences. She completed her PhD at DECIPHer, Cardiff University in 2021, and previously was employed by Health Data Research UK at Swansea University. Her research interests involve using quantitative research methods to answer important questions regarding child health and wellbeing. Her current research interests surround adverse childhood experiences (ACEs), the secondary harms of parental substance use, and vulnerable children (e.g. looked-after children). Within these, she is interested in the contribution of parenting, the family environment, and socioeconomic inequalities as potential modifying factors.

Dr. D. Rhian Barrance is a Lecturer at the School of Social Sciences at Cardiff University. Rhian is also Director of the WISERD Education Multi-Cohort Study, a longitudinal cohort study of secondary school children in Wales. This is a national study which is widely used by policymakers as a key source of evidence about children's experiences of growing up in Wales. Rhian's research interests include children's rights, assessment and equity in education. Much of her research focuses on the fairness of national assessments and qualifications in the UK from a children's rights perspective. She is also interested in children's social movements and environmental action and is currently working on a project on children and young people's involvement in social movements in Wales with colleagues at Cardiff, with a focus on children's activism around school policies and practices.

Dr. Charlotte Brookfield is a reader of Social Sciences and Co-Director of the Cardiff Q-Step Centre of Excellence in Quantitative Methods Teaching. The Cardiff centre is one of 15 centres based in UK higher education institutions exploring innovative and new approaches to engaging social science students with quantitative research methods. Her research interests are informed by the 'crisis' of number in social sciences and initiatives aimed at engaging social science students with quantitative approaches. Specifically, her recent research has focussed on the place of quantitative research in sociology.

References

Acar, S., and M. A. Runco. 2015. "Thinking in Multiple Directions: Hyperspace Categories in Divergent Thinking." *Psychology of Aesthetics, Creativity, and the Arts* 9 (1): 41–53. https://doi.org/10.1037/a0038501.

Amabile, T. M. 1996. *Creativity in Context: Update to the Social Psychology of Creativity*. Westview Press.

Aparicio, J., J. M. Cordero, and L. Ortiz. 2022. "Plausible Values and Their Use in Efficiency Analyses with Educational Data." *Applied Economics* 54 (29): 3340–3352. https://doi.org/10.1080/00036846.2021.2006136.

Avvisati, F., and F. Keslair. 2014. "'REPEST: Stata Module to Run Estimations with Weighted Replicate Samples and Plausible Values *Statistical Software Components S457918*." Boston College Department of Economics. revised 11 Dec 2024 (Accessed 24/06/2025).

Baer, J. 2011. "How Divergent Thinking Tests Mislead Us: Are the Torrance Tests Still Relevant in the 21st Century? The Division 10 Debate." *Psychology of Aesthetics, Creativity, and the Arts* 5 (4): 309–313. https://doi.org/10.1037/a0025210.

Baer, J. 2016. "Creativity Doesn't Develop in a Vacuum." *New Directions for Child and Adolescent Development* 151 (151): 9–20. https://doi.org/10.1002/cad.20151.

Barbot, B., and J. C. Kaufman. 2025. "PISA 2022 Creative Thinking Assessment: Opportunities, Challenges, and Cautions." *The Journal of Creative Behavior* 59 (1): e70003. https://doi.org/10.1002/jocb.70003.

Bernstein, B. 1971. *Class, Codes and Control*. Routledge and Kegan Paul.

Boden, M. A. 2004. *The Creative Mind: Myths and Mechanisms*. Routledge.

Boz, E. S. 2025. "Evaluating Measurement Invariance of students' Practices Regarding Online Information Questionnaire in PISA 2022: A Comparative Study Using MGCFA and Alignment Method." *Education and Information Technologies* 30 (1): 1219–1237. https://doi.org/10.1007/s10639-024-12921-7.

Brown, P., H. Lauder, and S. Y. Cheung. 2020. *The Death of Human Capital*. Oxford University Press.

Casey, C. 2012. *Economy, Work and Education*. Routledge.

Connolly, M. 2013. "The 'Liverpool Model(s)': Cultural Planning, Liverpool and Capital of Culture 2008." *International Journal of Cultural Policy* 19 (2): 162–181. https://doi.org/10.1080/10286632.2011.638982.

Connolly, M. 2025. *Creativity in Education, Urban and Cultural Policy: A Critique of a Contemporary Keyword*. Palgrave MacMillan.

Cresswell, J., and V. Plano Clark. 2017. *Designing and Conducting Mixed Methods Research*. Sage.

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

Csikszentmihalyi, M. 2013. *Creativity: The Psychology of Discovery and Invention*. Harper Collins.

de Vink, I. C., R. H. Willemsen, A. W. Lazonder, and E. H. Kroesbergen. 2022. "Creativity in Mathematics Performance: The Role of Divergent and Convergent Thinking." *British Journal of Educational Psychology* 92 (2): 484–501. https://doi.org/10.1111/bjep.12459.

Ericsson, K. A. 1999. "Creative Expertise as Superior Reproducible Performance: Innovative and Flexible Aspects of Expert Performance." *Psychological Inquiry* 10 (4): 329–333. https://doi.org/10.1207/S15327965PLI1004_5.

Fairclough, N. 2010. *Critical Discourse Analysis: The Critical Study of Language*. 2nd ed. Routledge. https://doi.org/10.4324/9781315834368.

Florida, R. L. 2002. *The Rise of the Creative Class : And How it's Transforming Work, Leisure, Community and Everyday Life*. Basic Books.

Glăveanu, V. P. 2013. "Rewriting the Language of Creativity: The Five A's Framework." *Review of General Psychology* 17 (1): 69–81. https://doi.org/10.1037/a0029528.

Glăveanu, V. P. 2015. "Creativity as a Sociocultural Act." *The Journal of Creative Behavior* 49 (3): 165–180. https://doi.org/10.1002/jocb.94.

Glăveanu, V. P., and Z. Sierra. 2015. "Creativity and Epistemologies of the South." *Culture & Psychology* 21 (3): 340–358. https://doi.org/10.1177/1354067X15601196.
Goecke, B., S. Weiss, and B. Barbot. 2025. "Content Validity of Creativity Self-Report Questionnaires from PISA 2022." *The Journal of Creative Behavior* 59 (2). https://doi.org/10.1002/jocb.70026.
Goldstein, H. 2017. "Measurement and evaluation issues with PISA." In *The PISA effect on global educational governanceThe PISA effect on global educational governance*, edited by V. Louis. New York and London: Routledge.
Grey, S., and P. Morris. 2018. "PISA: Multiple 'Truths' and Mediatised Global Governance." *Comparative Education* 54 (2): 109–131. https://doi.org/10.1080/03050068.2018.1425243.
Grey, S., and P. Morris. 2024. "Capturing the Spark: PISA, Twenty-First Century Skills and the Reconstruction of Creativity." *Globalisation, Societies and Education* 22 (2): 156–171. https://doi.org/10.1080/14767724.2022.2100981.
Guo, Y., S. Lin, Z. J. Williams, T. C. Grantham, J. Guo, L. Q. Cole Clark, and W. Zou. 2024. "Creative Potential and Creative self-Belief: Measurement Invariance in cross-Cultural Contexts." *The Journal of Creative Behavior* 58 (2): 209–226. https://doi.org/10.1002/jocb.643.
Kagan, S., and D. Dumas. 2025. "More Creative Activities, Lower Creative Ability: Exploring an Unexpected PISA Finding." *The Journal of Creative Behavior* 59 (2). https://doi.org/10.1002/jocb.70035.
Kant, I. 1790/2013. *Critique of the Power of Judgment*. Cambridge University Press.
Karwowski, M. 2014. "Creative Mindsets: Measurement, Correlates, Consequences." *Psychology of Aesthetics, Creativity, and the Arts* 8 (1): 62–70. https://doi.org/10.1037/a0034898.
Kaufman, J. C., M. L. Evans, and J. Baer. 2010. "The American Idol Effect: Are Students Good Judges of Their Creativity Across Domains?" *Empirical Studies of the Arts* 28 (1): 3–17. https://doi.org/10.2190/EM.28.1.b.
Kim, K. H., and S. G. Park. 2020. "Relationship Between Parents' Cultural Values and Children's Creativity." *Creativity Research Journal* 32 (3): 259–273. https://doi.org/10.1080/10400419.2020.1821566.
Klausen, S. H. 2010. "The Notion of Creativity Revisited: A Philosophical Perspective on Creativity Research." *Creativity Research Journal* 22 (4): 347–360. https://doi.org/10.1080/10400419.2010.523390.
Long, H., B. A. Kerr, T. E. Emler, and M. Birdnow. 2022. "A Critical Review of Assessments of Creativity in Education." *Review of Research in Education* 46 (1): 288–323. https://doi.org/10.3102/0091732x221084326.
Lubart, T., V. P. Glăveanu, H. de Vries, A. Camargo, and M. Storme. 2019. "Cultural Perspectives on Creativity." In *The Cambridge Handbook of Creativity*, edited by J. C. Kaufman and R. J. Sternberg, 421–447. Cambridge University Press.
Lucas, B., G. Claxton, and E. Spencer. 2013. "Progression in Student Creativity in School: First Steps Towards New Forms of Formative Assessments." *OECD Education Working Papers, No. 86*. OECD Publishing. http://dx.doi.org/10.1787/5k4dp59msdwk-en.
Muthén, L., and B. Muthén. 2017. "Mplus: User's Guide." Muthén & Muthén. Accessed December 18, 2025. https://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf. Eighth ed.
Negus, K., and M. Pickering. 2004. *Creativity, Communication and Cultural Value*. Sage Publications.
O'Connor, J. 2024. *Culture Is Not an Industry: Reclaiming Art and Culture for the Common Good*. Manchester University Press.
OECD. 2019a. *Creative Thinking Framework* (Third Draft). OECD. OECD Publishing Paris. Available At: PISA 2021 Creative Thinking Framework (Third Draft). Accessed December 18, 2025.
OECD. 2019b. "Future of Education and Skills 2030. Concept Note. Conceptual Learning Framework. Learning Compass 2030." OECD Publishing. Accessed December 18, 2025. www.oecd.org/education/2030-project/learning.

OECD. 2022. *Thinking Outside the Box: The PISA 2022 Creative Thinking Assessment*. OECD Publishing. PISA 2022 Creative Thinking | OECD. Accessed December 18, 2025.
OECD. 2024a. *PISA 2022 Technical Report*. OECD Publishing Paris. PISA 2022 Technical Report | OECD. Accessed December 18, 2025.
OECD. 2024b. "New PISA Results on Creative Thinking: Can Students Think Outside the Box? OECD." OECD Publishing Paris. Accessed December 18, 2025. https://www.oecd.org/en/publications/new-pisa-results-on-creative-thinking_b3a46696-en.html.
OECD. 2024c. "PISA 2022 Results (Volume III) Creative Minds Creative Schools." Accessed December 18, 2025. https://www.oecd.org/en/publications/pisa-2022-results-volume-iii_765ee8c2-en.html.
OECD. 2024d. "Social and Emotional Skills for Better Lives: Findings from the OECD Survey on Social and Emotional Skills 2023." OECD Publishing Paris. Accessed December 18, 2025. https://doi.org/10.1787/35ca7b7c-en.
Pappano, L. 2014. "Creativity Becomes an Academic Discipline." *The New York Times*. Accessed February 5, 2014.
Paul, E. S., and S. B. Kaufman. 2014. *The Philosophy of Creativity: New Essays*. Oxford University Press.
Paul, E. S., D. Stokes, E. N. Zalta, and U. Nodelman. 2024 "The Stanford Encyclopedia of Philosophy." Accessed December 18, 2025. https://plato.stanford.edu/archives/spr2024/entries/creativity/ .
Pirrie, A. 2016. "The LEGO Story: Remolding Education Policy and Practice." *Educational Review* 69 (3): 271–284. https://doi.org/10.1080/00131911.2016.1207614.
Plucker, J. A., R. A. Beghetto, and G. T. Dow. 2004. "Why Isn't Creativity More Important to Educational Psychologists? Potentials, Pitfalls, and Future Directions in Creativity Research." *Educational Psychologist* 39 (2): 83–96. https://doi.org/10.1207/s15326985ep3902_1.
Robinson, K. 2011. *Out of Our Minds Learning to Be Creative*. Capstone.
Runco, M. A., G. Millar, S. Acar, and B. Cramond. 2010. "Torrance Tests of Creative Thinking as Predictors of Personal and Public Achievement: A Fifty-Year Follow-Up." *Creativity Research Journal* 22 (4): 361–368. https://doi.org/10.1080/10400419.2010.523393.
Runco, M. A., J. A. Plucker, and W. Lim. 2001. "Development and Psychometric Integrity of a Measure of Ideational Behavior." *Creativity Research Journal* 13 (3–4): 393–400. https://doi.org/10.1207/S15326934CRJ1334_16.
Said-Metwaly, S. K., W. Van den Noortgate, and E. Kyndt. 2017. "Approaches to Measuring Creativity: A Systematic Literature Review." *Creativity Theories-Research-Application* 4 (2): 238–275. https://doi.org/10.1515/ctra-2017-0013.
Schleicher, A. 2019. *OECD Learning Compass 2030*. OECD Learning Compass 2030, Vancouver Canada. Available at: Bing Videos Accessed December 18, 2025.
Schleicher, A. 2024. "Introduction to PISA Creativity Test Results. Available at: Andreas Schleicher Presents PISA 2022 Volume III - Creative Thinking - 18 June 2024 PPTX." Accessed December 18, 2025.
StataCorp. 2017. *Stata Statistical Software: Release 15* [Computer Software]. StataCorp LLC.
Stein, M. I. 1953. "Creativity and Culture." *The Journal of Psychology* 36 (2): 311–322. https://doi.org/10.1080/00223980.1953.9712897.
Sternberg, R. J., V. Glăveanu, and J. C. Kaufman. 2024. "In Quest of Creativity: Three Paths Toward an Elusive Grail." *Creativity Research Journal* 36 (1): 155–175. https://doi.org/10.1080/10400419.2022.2107299.
Vincent-Lancrin, S. 2019. *Fostering Students' Creativity and Critical Thinking: What it Means in School, Educational Research and Innovation*. Paris: OECD Publishing. https://doi.org/10.1787/62212c37-en.
Weisberg, R. W. 2006. "Modes of Expertise in Creative Thinking: Evidence from Case Studies." In *The Cambridge Handbook of Expertise and Expert Performance*, edited by K. A. Ericsson, 761–788. Cambridge: Cambridge University Press.

Wong, R., and W. Niu. 2013. "Cultural Difference in Stereotype Perceptions and Performances in Nonverbal Deductive Reasoning and Creativity." *The Journal of Creative Behavior* 47 (1): 41–59. https://doi.org/10.1002/jocb.22.

Xiaomin, L., and E. Auld. 2020. "A Historical Perspective on the OECD's 'Humanitarian turn': PISA for Development and the Learning Framework 2030." *Comparative Education* 56 (4): 503–521. https://doi.org/10.1080/03050068.2020.1781397.

Young, M. F. D. 2008. *Bringing Knowledge Back In: From Social Constructivism to Social Realism in the Sociology of Education*. Routledge.