The Influence of School Evaluation Expectation on Adolescent Creative Potential

Jean Urbaniak[1]

Georgios Valmas[1]

Baptiste Barbot[1,2]

[1] UCLouvain (University of Louvain), Louvain-la-Neuve, Belgium

[2] Child Study Center, Yale University, New Haven, CT, USA

Author Note:

Baptiste Barbot, Psychological Sciences Research Institute, UCLouvain & Child Study Center, Yale University. Correspondence to Baptiste Barbot, Psychological Sciences Research Institute, UCLouvain, Place Cardinal Mercier 10/L3.05.01, 1348 Louvain-la-Neuve, Belgium. Email: baptiste.barbot@uclouvain.be. ORCID: https://orcid.org/0000-0002-5096-2596

Conflict of Interest: The authors declare that they have no conflict of interest.

Working paper

Abstract

Worldwide efforts are underway to bolster creativity in school settings. Hence, there is a critical need to accurately assess creativity in this setting to monitor the effectiveness of educational policies and programs aiming to foster students' creativity. However, there is a lack of research on the effect of variables unique to the school context that may hinder creative performance. This study investigated how expected school evaluation influences adolescents' creative potential. 428 Belgian students from the first to the fifth year of secondary school completed four creativity tasks (varying in terms of modalities and thinking processes involved). The explicit evaluation expectation was experimentally manipulated. Findings demonstrated its overall detrimental effect on several aspects of creative potential, with a differential effect according to the creative task considered. Some of these effects were mitigated by school grade and stress-state level. Results are discussed in light of the evaluative context literature and future directions for creativity assessment in school settings.

Educational relevance statement

This study demonstrates that the school evaluation context can be detrimental to performance in specific components of creative potential (i.e., specifically associative thinking and fluency in divergent thinking). This suggests that school evaluations involving creative responses should be cautiously designed to prevent response biases. In that regard, teachers are encouraged to foster a *low evaluation stakes* climate during creative activities, characterized by supportive guidance and no focus on grades.

The study also revealed that the impact of the evaluation context varies according to school grade, with students in certain pivotal years being more vulnerable to evaluative context. In this study, the third year of Belgian secondary school (third year of ISCED level 2) was particularly sensitive, a period that align with the *seventh-grade slump* — a time of temporary decline of divergent thinking. Although the timing of this phenomenon may vary by country, particular

attention should be placed on providing an environment free of evaluation stakes during this period of vulnerability.

Finally, findings suggest that students' stress state moderates the effects of evaluation context on associative thinking: under evaluative pressure, students with lower stress levels underperformed compared to those with higher stress levels, who may have been more motivated to perform. These results may inspire further studies examining the importance of the emotional state in task engagement at school.

Keywords: creativity, evaluation context, adolescence, stress state.

The Influence of School Evaluation Expectation on Adolescent Creative Potential

1. Introduction

1.1. Context

Worldwide efforts are underway to tailor education to 21st-century skills, including creativity (Plucker et al., 2015). General interest in making creativity a more central goal in schools has increased over the years (Kaufman et al., 2016), and many countries have targeted creativity as a key skill to foster in their educational system (Stables, 2009). This is notably the case in Wallonia, Belgium (where this study took place), where creativity has been identified as one of the seven transversal skill areas targeted by the "Pact for Excellence in Teaching" ["*Pacte pour un Enseignement d'excellence*"] (Fédération Wallonie-Bruxelles, 2023). Illustrating this increasing international interest in creativity in education, the Programme on International Student Assessment (PISA) measured the creative thinking skills of 15-year-old adolescents for the first time in 2022, assessing their ability to engage productively in the generation, evaluation, and improvement of ideas (OECD, 2024).

However, to develop creative skills at school, there is a need to understand how creativity develops and what factors influence this development, particularly in school settings. Relatedly, there is a critical need for accurate creativity assessment (Reiter-Palmon et al., 2019), to monitor the effectiveness of educational efforts (Lubart & Besançon, 2017) and to support and guide students' creativity development (Barbot, 2019). Although creativity has been assessed for decades in school settings for research or talent identification purpose (Kaufman et al., 2016), the three primary goals of assessment in education have not been fulfilled: assessment to support learning (formative assessment), assessment for certification, progress, and transfer (summative assessment), and assessment for accountability (e.g., to evaluate programs' efficiency) (Archer, 2017). In sum, although creativity is increasingly promoted as a key competence in education policies, its assessment in schools remains a complex and often neglected issue.

In actual school practice, teachers' evaluation of students' creativity often relies on subjective perceptions, shaped by their own implicit theories of creativity, rather than on structured assessments, which leads to limited recognition of students' creative potential (Gralewski & Karwowski, 2018). Creativity assessment is typically confined to specific disciplines (e.g., arts education) or isolated activities (e.g., creative writing, mathematical problem solving). Nevertheless, new standardized tools are being developed to address the increasing need for valid creativity assessments in school settings (e.g., Cebeci & Acar, 2025). Despite these efforts, creativity assessment should, however, account for the multiple mitigating factors of creative performance unique to the school environment (Rubenstein et al., 2018).

Regrettably, there is a lack of research on the effect of variables unique to this context that may hinder creative performance. In particular, the effect of *expected evaluation* in a school-specific context and their possible interaction with creativity tasks' constraints seem overlooked. Another aspect that must be considered is the *level of evaluation stakes*: Is the evaluation 'low stakes,' in that evaluation is only formative and has little consequences on students' progression, or is the evaluation summative, with 'high stakes' consequences on an end-of-year pass or fail decision? The response to this question could explain variations in the impact evaluation has on creative performance. Yet, the effect of the evaluation context on creativity is overlooked, and the few studies that have investigated it have shown contradictory results (Yuan & Zhou, 2008). Furthermore, these studies often lack ecological validity in that they were not conducted in the school context. Hence, the present study examines how the evaluation expectation context (i.e., *expected evaluation context* vs *practice context*) and the level of evaluation stakes (i.e., *low stakes* vs *high stakes*) influence adolescents' creative potential in a range of creative tasks (varying in terms of modalities and thinking processes involved).

1.2.Creativity in educational settings

Navigating creativity in educational settings is a complex undertaking, requiring every stakeholder (educators, teachers, researchers, psychologists) to understand the concept. Even if its definition is controversial (Acar et al., 2017), creativity is often defined as the ability to bring a novel and appropriate solution to a given problem (e.g., Runco & Jaeger, 2012). However, creativity is multifaceted, and not a "generic" or unitary construct (Barbot, 2018a). In educational studies of creativity, a distinction is generally made between different aspects of the creativity phenomenon, including creative achievements (i.e., observable products or effectively applied ideas, Karwowski et al., 2021), creative activity or participation (e.g., Nikkola et al., 2024), and above all creative potential. In line with componential or multivariate approaches to creativity (e.g., Sternberg & Lubart, 2004), creative potential is the set of individual (cognitive, conative, emotional) and environmental resources that contribute to creative success (e.g., Lubart & Guignard, 2004). Among these resources, one of the most studied aspects is divergent thinking (i.e., the ability to produce many creative ideas in response to a single stimulus or problem; e.g., Reiter-Palmon et al., 2019). Divergent thinking (DT) does not equate to the whole creativity phenomenon, but it is known to be a central component of creative ideation (Runco & Acar, 2012) despite its limited power to predict real life creative achievements (Said-Metwaly et al., 2022) . Another central cognitive component of creative potential is associative thinking, which is the capacity to link concepts from memory (Beaty & Kenett, 2023). Associative abilities can be efficiently and reliably assessed, making them valuable as fundamental indicators of creative potential (Benedek et al., 2012).

Beyond the multiple facets of creativity, its overall development is another key issue in understanding creativity in school settings. Indeed, the primary goal of education is to foster the development of skills, and creativity is no exception. In this regard, research indicates that it does not develop linearly throughout the lifespan (Barbot & Rogh, 2020), with different

developmental trajectories regarding the dimensions and criteria considered (Gralewski et al., 2016). Nevertheless, DT has been the primary focus of most developmental studies, largely because it can be assessed from a very early age. Consequently, much of our understanding of creativity development is derived from knowledge on DT (Barbot & Rogh, 2020). A meta-analysis of DT development suggests that even if the general developmental trend is an increase of DT during school age, a developmental slow-down (or "slump") is observed around the age of 9-10 (fourth grade) as well as a more marked slump around the age of 12-13 (seventh grade) (Said-Metwaly et al., 2021). This meta-analysis outlined considerable inter-study variance as studies are generally inconsistent regarding the period of appearance of slumps, their intensity, and duration (see Barbot et al., 2016). To understand this developmental phenomenon, a range of explanations have been proposed, including a historical interpretation (Torrance, 1968) according to which the school environment can be detrimental as it may trigger social conformity or generate stress related to the transition to high school (Barbot & Lubart, 2012). Indeed, the environment enables individuals to mobilize their cognitive resources, arouses emotions, and exacerbates or inhibits some conative factors such as motivation. However, the impact of other school-environment related variables (such as evaluation context) has not been closely examined.

The school environment, where students spend most of their time, thus critically influences students' creativity development. Illustrating one of many creative contradictions in education (Beghetto & Sriraman, 2017), a prevalent idea suggests that school "kills creativity". Although little data supports this thesis (Barbot & Said-Metwaly, 2021; Karwowski, 2022), a body of research indicates that the school environment is not "neutral" in many ways (Rubenstein et al., 2018).

In this regard, several studies indicate a negative impact of the school evaluation context on students' creative potential. For example, Amabile (1979) and Hu et al. (2018) highlighted

that students expecting external evaluations were less creative than those who did not expect such evaluations for artistic tasks. Bartis et al. (1988) found the same findings with students performing an Alternative Uses Task (AUT), as well as Amabile et al. (1990), both for a verbal and an artistic task, suggesting that the effect of evaluation expectation is not limited to just one type of task. Along the same lines, Silvia and Phillips (2004) concluded that increasing self-evaluation had a detrimental effect on performance in generating remote associates and finding unusual uses when people did not feel able to improve. Building on the work of Duval and Wicklund (1972), according to which expecting external evaluation increases self-evaluation, this finding aligns with previous work arguing for a detrimental effect of the evaluation context on creativity. Baer (1997) added that this effect was moderated by gender, with a stronger impact on girls.

Nevertheless, some findings did not identify a similar pattern and found more balanced findings. For example, Yuan and Zhou (2008) showed a negative impact of expectation of external evaluation on the blind-variation process (i.e., in which a variety of ideas are generated) but a positive effect of such expectation on the selective-retention process (i.e., in which ideas are selected and modified) to solve a human resource management problem creatively. These findings suggest that the impact of expected external evaluation on creativity is task-sensitive, depending on the extent to which the creative product must have a novelty dimension (i.e., requiring more blind-variation skills) or be appropriate to a specific context (i.e., emphasizing the selective-retention process; Yuan & Zhou, 2008). The results of Camarda et al. (2021) are also mitigated. According to the authors, the impact of direct supervision of an adult expert on creative ideation depends on age, with detrimental effects on children, beneficial effects on adolescents, and no effects on late adolescents. According to Xue et al. (2020), originality, fluency, and flexibility were lower in young adolescents under expected evaluation in a scientific

DT task, but artistic creativity was boosted under such conditions. Finally, Miller (2007) found no main effect of expected evaluation on artistic creativity.

Some studies even conclude that the evaluation context globally has a favorable impact on creative performance. For example, Shalley (1995) found that DT was stimulated by expected evaluation among adults solving a human resource problem. Shalley and Perry-Smith (2001) outline that creativity level is higher when people anticipate an informational (i.e., providing information to improve performance) rather than a controlling (i.e., gauging how well one performs relative to a set standard) evaluation. Moreover, Eisenberger and Rhoades (2001) report higher creative performance when a financial reward is promised to children, students, and adults who are performing well according to judges. Finally, Lei et al. (2021) conclude that adolescents were more original when performing a DT task expecting evaluation.

Despite these contradictory findings, there are indeed theoretical reasons to assume that the evaluation context influences creative potential. Several factors known to affect creativity could be impacted by such expectations, including personality traits (Jirásek & Sudzina, 2020) such as perseverance (Sternberg & Lubart, 1995), motivation (Bonnardel, 2009), self-esteem (Barbot, 2020), or creative self-beliefs (Karwowski et al., 2019). In particular, motivation could be the most impacted component of creative potential under expected evaluation. Intrinsic motivation is known to be positively associated with creativity (e.g., de Jesus et al., 2013), while extrinsic motivation is considered detrimental to creativity (Amabile, 2011). However, the relationship between extrinsic motivation and creativity is more complex: rewards can benefit creativity under certain conditions. First, extrinsic motivators can enhance creativity by increasing intrinsic motivation when they give information that confirms people's competencies or gives value to their work (Amabile & Pratt, 2016). Social gratification or good grades arising from school evaluation could fall within this category of extrinsic motivators. However, if students are mainly extrinsically motivated to complete a creative task, the evaluation context

could be problematic as they may apply the “rat race for grades” (Pannier, 2005) and likely focus more of their efforts on evaluation criteria (e.g., sticking more closely to task instructions) than on their creative outputs. On that line, it is thus possible that the evaluation context also interacts with the established differential effects of task instructions on creative performance (e.g., "be creative" vs. "be fluent" effect; Reiter-Palmon et al., 2019). According to self-determination theory (Ryan & Deci, 2000), different types of extrinsic motivation can arguably be triggered in a school context. When evaluation is expected, lower self-determined forms of extrinsic motivation, such as external regulation (based on rewards and punishments) and introjected regulation (driven by guilt avoidance and self-esteem protection), may be favored over more autonomous forms, such as identified regulation (based on recognizing the value of the task) and integrated regulation (stemming from a congruence between personal and task-related values). However, practice-oriented contexts, which are often associated with lower perceived stakes and fewer performance-related consequences, may foster more autonomous forms of motivation by reducing external pressure (Lepper et al., 1973). This highlights the importance of considering the stakes underlying external evaluation in a school context, which can impact students’ engagement (Borger et al., 2025) and creative performance (Kruglanski et al., 1971).

In sum, the impact of the school context, especially the school evaluation context, on creative performances appears manifold. The evaluation expectation impacts many environmental, conative, cognitive, and emotional factors known to be linked with creative performance. However, little is known about the direct influence of this evaluation context (and evaluation expectation) on creative performance, as this factor is often overlooked in creativity studies that take place in school settings, and the few studies that investigated it provided contradictory results. This line of research deserves more attention as a prerequisite for an accurate evaluation of creativity in school settings.

1.3. Present study

The impact of evaluation context on creativity does not seem well established in extent research literature. Yuan and Zhou (2008) state that contradictory results from earlier studies stem from a lack of understanding of the cognitive processes involved in creative ideation. However, other factors could explain the diverging results obtained by different studies. Indeed, most studies concluding a detrimental effect of expected evaluation on creative performances were conducted with children, while most studies adding nuance to these results were conducted among young adults. Thus, it is reasonable to question the possible moderating effect of age on the effects of expected evaluation on creative performance. Generally, studies have not shown enough interest in possible moderating variables that could impact the relationship between evaluation context and creative performance, such as age and gender. Another explanation for the inconsistency of findings in this line of work is considering the specific tasks used in studies, as creative performance is somewhat task-specific (e.g., Reiter-Palmon et al., 2009). Furthermore, in former studies, evaluation expectation has been manipulated in various ways: through the supervision of an adult (e.g., Camarda et al., 2021), by an expected reward (e.g., Eisenberger & Rhoades, 2001), by expected grading (e.g., Yuan & Zhou, 2008), or by increasing self-awareness and self-evaluation (e.g., Silvia & Phillips, 2004). These diverse manipulations reflect a lack of consensus on how best to conceptualize the evaluation context. However, they can all be meaningfully related to standard classroom practices. For instance, teacher supervision is systematically present in evaluative settings to prevent cheating; good grades serve as extrinsic rewards for many students; and even the simple act of writing one's name on a test sheet can heighten self-awareness and trigger self-evaluation processes. Acknowledging this conceptual ambiguity, the present study reconciles these different conceptualizations within a school-specific framework of evaluation context.

The objective of this study was, therefore, to examine how school evaluation context and evaluation stakes influences adolescents' creative potential in a range of tasks (varying in terms of modalities and thinking processes involved) while considering a set of variables known to impact creative potential, including age and gender (Said-Metwaly et al., 2021; Taylor et al., 2024), and variables more specific to school context including school year (school year transitions being suspected to perturb creative abilities; Torrance, 1968) and stress state (known to hinder creative potential; Byron & Khazanchi, 2011). In such a context, we formulated three hypotheses:

H1: We hypothesized that expected school evaluation would have a detrimental main effect on both divergent and associative thinking skills, consistent with prior research suggesting that external evaluation undermines creative performance (e.g., Amabile, 1979; Hu et al., 2018).

H2: We expected that school year would moderate this effect, in line with Torrance's (1968) hypothesis that the seventh-grade slump partly results from an increased fear of social judgment, suggesting that age may shape the relationship between social evaluation and creativity.

H3: We hypothesized that stress level would moderate the effect of expected evaluation on creative potential, based on theories suggesting that stress contributes to the detrimental impact of school environments, an effect that may be exacerbated in evaluative contexts (Barbot & Lubart, 2012).

While creativity is a multifaceted construct, this study focuses specifically on divergent and associative thinking, as they represent two core and complementary cognitive components of creative potential that are both theoretically grounded and measurable through standardized tasks.

2. Materials and methods

This study consists of a quasi-experimental quantitative research in a naturalistic context. To address our research objective, we administered four creative tasks targeting crucial creative potential components that are associative thinking (i.e., visual and verbal associative thinking tasks) and DT (i.e., visual and verbal creative ideation tasks) across visual/graphic and verbal/written modalities, to different cohorts of students divided into two groups distinguished by the test conditions (*explicit evaluation expectation context*).

2.1. Participants

Participants were 428 Belgian students (mean age = 14.31; SD age = 1.58; 51% female), from 12 to 19 years schooled in 1st to 5th year of secondary school (i.e., from the first year of the International Standard Classification of Education (ISCED) level 2 to the second year of ISCED level 3). For each school year, four classes from a large, middle-income secondary school[1] in Wallonia region, were randomly selected.

2.2. Measures

2.2.1. Visual creative ideation

The incomplete shapes task (Barbot, 2018b) is a graphic creative ideation task aimed at turning abstract shapes into original drawings using a tablet-based interface. Unlike most DT graphic tasks that use the same stimulus to generate as many responses as possible, the incomplete shapes task uses the MTCI (Multi Trial Creative Ideation) paradigm (Barbot, 2018b) to propose a structured set of 7 different stimuli, hence offering the advantage of limiting stimulus-dependency (due to the use of multiple stimuli) and fluency-dependency (as the number

[1] This school has a socio-economic index of 12 out of 20 (Fédération Wallonie-Bruxelles, 2020), based on parents' income, their level of education, their unemployment rate, activity rate, and percentage of social aid beneficiaries, as well as their professional activity (Fédération Wallonie-Bruxelles, 2024), placing it in the Walloon average.

of responses is standardized). Participants must generate a unique original idea for each stimulus without time limit but with time recorded in the background. The instruction did not mention that the task had to be quickly completed. For this task, fluency, which is the most commonly used dimension of creative ideation in creativity in research (Reiter-Palmon et al., 2019), is scored in two different ways.

Firstly, the average "exploration time" (i.e., the time between the appearance of the visual stimulus and the start of the drawing activity, signaling that an idea has been formulated) serves as a proxy for ideational fluency. This approach is consistent with the original conceptualization of Guilford (1950; i.e., rate of ideation per unit of time, or response speed) and previous research (Barbot, 2018b), where shorter exploration times reflect a higher fluency in formulating creative responses. The rationale is that delayed response initiation may indicate greater cognitive effort or difficulty in idea generation. Furthermore, operationalizing fluency through response time relaxes the constraints of time-limited tasks, which is an essential consideration given that time pressure can negatively impact the quality of responses in creative ideation tasks (Forthmann et al., 2020). It also provides more room for persistence and sustained effort, which are essential components in the production of original ideas (Nijstad et al., 2010). This metric is particularly relevant given that traditional fluency measures (i.e., number of ideas generated) are constrained by the MTCI paradigm.

Secondly, a classic ideational fluency equivalent is calculated by considering the total production time, resembling fluency scores in "gold standard" DT measures (e.g., Torrance, 2008: TTCT). As fluency alone is a limited predictor of real-life creative accomplishments (Plucker et al., 2011), we also derived a quality indicator of responses (Reiter-Palmon et al. 2019). Specifically, a creativity score was generated with AuDrA (i.e., Automated Drawing Assessment for evaluating creativity; Patterson et al., 2024), a machine learning model trained on a large dataset of human-rated responses to the Incomplete Shapes task. The model predicts

human-rated creativity scores with high accuracy (i.e., $r > .90$; Patterson et al., 2024), offering a scalable and objective alternative to subjective expert ratings. Creativity, in this context, is defined as the perceived originality of responses, aligning with Amabile's Consensual Assessment Technique (CAT; Amabile, 1982). Finally, the "black ink" method, operationalized as the proportion of dark pixels on a white background, was also used to automatically generate an elaboration score which is both objective and efficient (e.g., Acar et al., 2024).

2.2.2. Verbal creative ideation

The Alternative Uses Task (AUT) is a classic verbal DT task that involves generating a maximum of alternative uses for a given object within a limited time (Guilford, 1967). In this research, participants were asked to imagine as many unusual ways to employ a *hat* as possible within 120 seconds. To score this task, we used a traditional approach (Reiter-Palmon et al., 2019), generating a fluency and an originality score, the most commonly reported indicators of quantity and quality in DT assessment (Acar & Runco, 2017; Saretzki et al., 2024). The fluency score corresponds to the number of given appropriate answers and is the primary historical measure in DT tests (Silvia et al., 2013). For each response, an originality score was generated using Open Creativity Scoring with Artificial Intelligence (OCSAI; Organisciak et al., 2022), which leverages advanced language models to enhance the scoring of creative tasks related to originality and divergent thinking. Compared to other methods, OCSAI demonstrates a significantly higher correlation with human judgments (up to $r = .813$), greatly outperforming previous approaches like semantic distance scoring ($r = .256$ for the Open Creativity Scoring system; Dumas et al., 2021) in assessing the originality of ideas. Finally, two individual originality variables were computed: the first is based on the average originality score across all responses, while the second corresponds to the score of the most original idea (*top-scoring method*, e.g., Benedek et al., 2013). Flexibility was not assessed as a separate score, as response

categories from which flexibility scores are derived are often arbitrary (Reiter-Palmon et al., 2019), and as it currently lacks scalable automated scoring methods.

2.2.3. Visual associative thinking

The Pictures association task (PAT) [masked for peer-review] is a visual associative thinking task in which participants must associate images to a target image according to their own criteria. The PAT builds upon classic objects categorization paradigms and procedures such as the Alternative Categories Task (ACT, Barsalou, 1983) or the Goldstein Object Sorting Test (see Payne & Friedlander, 1962). Specifically, the task starts with an identification phase, whereby participants review 78 "solution" images used in the task, and they must select those images depicting objects that they have never seen before. Next, for each of the three items of the task, a target image acting as a stimulus is shown to the participants, who can select among the 78 solution images those that can be linked to the target image. Consistent with the three stimulus items of the task (a taxicab, a chicken, and a kitchen pot), the images belong to three conceptual categories, including transportation, farm, and kitchen items, and also include three "non-objects" pictures used for response quality control. However, as in classic associative thinking measures (e.g., word association tasks), there are no right or wrong answers, as participants may see links between images that are not obvious. Hence, the same solution image (e.g., picture of a horse) may be associated with target items of distinct categories (e.g., taxicab and chicken), reflecting the flexible activation and overlap of semantic networks. While the selection of non-objects in the identification phase supports the validity of the protocols (i.e., non-objects are not supposed to be known by participants and may not be used as associates in the association phases), the number of items selected in this phase allows for adjusting the final associative score to account for differences in respondents' knowledge of the depicted objects. Although several associative thinking metrics are derived from this task, this study used only the total visual associative thinking score, computed as the average number of associations made

across all three items. Evidence of criterion validity of this score is supported in recent research showing its sizeable relationship with creative writing performance (Golab & Barbot, 2024). In this study, the composite associative thinking score showed good internal consistency (α = 0.829).

2.2.4. Verbal associative thinking

The word association task is a classic verbal associative thinking paradigm that involves writing a maximum of words related to a target word within a limited time (Benedek et al., 2012; Taylor & Barbot, 2021). This paradigm is commonly used in associative thinking research and has shown strong relationships with creative thinking and performance (e.g., Beaty & Kenett, 2023; Benedek et al., 2012). For this research, the words "*apple"* and "*paper*" were used as stimuli, and the participants had 120 seconds to complete each stimulus. Each response is counted as one idea, and the total number of valid entries constitutes the fluency score. A response was considered valid if it was a real word and judged by the first two authors as semantically meaningful in relation to the target word. A composite associative thinking score was computed by aggregating the scores (i.e., the number of valid responses) from the two items. The intra-class correlation (ICC) between items showed a satisfactory internal consistency (average measures = 0.750, 95% CI [0.697, 0.794]).[2]

[2] Initially, two originality scores were added to the composite associative thinking score for this task. They were based on the average and the highest originality score of the responses and were generated using OCSAI (Organisciak et al., 2022).These scores couldn't be aggregated given the ICC for average measures was 0.176 (95% CI [-0.009, 0.326]) for the average originality score, and 0.260 (95% CI [0.094, 0.395])) for the highest score. So, those indexes were excluded in our analyses.

2.2.5. Other measures

For this study, a stress state measure based on the Cognitive Test Anxiety Scale (Cassady & Johnson, 2012) was adapted and translated into French. This newly adapted scale comprised five items (e.g., I felt under a lot of pressure to perform well on this test) and yielded reliable scores with acceptable internal consistency ($\alpha = 0,756$). Additionally, three items questioning the digital proficiency level (e.g., "I felt comfortable using the tablet") were added to the questionnaire and used for statistical control purposes. The internal consistency of this scale was high ($\alpha = 0,860$).

2.3. Procedure

After securing parental/guardians' (passive) consent via an information email, participants were invited to sign an informed consent before the start of the study. In order to minimize response biases, this consent form provided limited information about the specific aim of the study, describing research on creativity development without mentioning our interest in evaluation expectation effects. A debriefing, including more information about the study's ultimate goals, was provided after completion. The ethical commission of the [masked for peer-review] approved the research protocol. The study occurred within the participants' school without altering the classroom groups and during regular class hours to keep the testing conditions as close as possible to a school evaluation. Participants completed the measures in classroom settings using digital tablets, which can unnoticeably record process data without impacting creative performances (Barbot, 2018b). To ensure consistency, the experimenter was always present in the classroom and personally delivered the standardized instructions for both the evaluation and practice contexts. Building upon an experimental design inspired by Yuan and Zhou (2008), evaluation expectation (*explicit evaluation expectation context*) was manipulated by instructing half of the participating classrooms that the results of their tests would be communicated to their teachers and considered during class councils (*expected evaluation* modality). The other half were instructed to complete tasks for practice purposes (i.e., *practice* modality, acting as control). More precisely, for the *practice* modality, students were informed that the first two tasks were for practice and would not

be evaluated (phase 1), serving as preparation for the "real" test composed of the last two tasks (phase 2). Only the first part of the test was considered as a practice context.

Consistent with Yuan and Zhou (2008), counter-balancing of tasks order was implemented as follows: for each school year, two classes performed the incomplete shapes task and the word association task in Phase 1, while they performed the alternative uses task and pictures association task in Phase 2 ("*order A*"). One of these classes underwent the *practice* modality in Phase 1, while the other performed Phase 1 in the *expected evaluation* context. The other two classes performed the creative tasks in reverse order ("*order B*"), with one class experiencing the *practice* modality and the other the expected evaluation modality. Between the two phases, all students were asked to complete a questionnaire to measure their stress state during Phase 1 and their digital proficiency level.

2.4. Analytic plan

To address our research questions, we ran several MANCOVA analyses to study the effect of explicit evaluation expectation context and its interaction with school year on multiple creativity scores from the selected tasks. This approach allowed for the statistical control of the effect of variables known to be related to creativity (e.g., age, gender) as well as for methodological variables (e.g., phase, tablet level proficiency). This approach was preferred over a multilevel analysis given that all tested students were sampled in the same school, with predominantly the same professors across the classes (thus eliminating any significant "teacher effect" between classes).

3. Results

3.1. Preliminary analyses

First, univariate outliers were detected (i.e., values that deviate from the mean by at least two standard deviations, Cohen et al., 2003) and treated using a winsorization method. After detection of non-normality across several variables of interest (i.e., fluency and elaboration

scores of the visual creative ideation task and visual associative thinking score), we applied the Rankit transformation to normalize their distribution and satisfy the necessary statistical assumptions for planned analyses.

Data screening indicated that a small proportion of data was missing across the different creative tasks (5.0% on average, ranging from 2.3% for IST to 13.6% for AUT). Missing data primarily resulted from technical issues during the computer-based administration of the tasks (e.g., incomplete recordings or corrupted files). Little's MCAR was non-significant, $\chi^2(103)= 124.6$, p= .08, suggesting that the missingness could be considered missing completely at random (MCAR). Given this pattern, listwise deletion was applied. As a result, the effective sample size varied depending on the analysis, though results could reasonably generalize to the complete sample.

Two preliminary MANCOVAs (one per task set, i.e., order A or order B) were ran to ensure that the control groups were comparable to the experimental groups in terms of initial level of performance in the creativity tasks, controlling for digital proficiency level, phase, gender, and age. Results showed no significant differences between the groups for the tasks set in Order A ($F(3,257)= .388$, $p = .746$) and in Order B ($F(7,276)= 1.207$, $p= .299$). Hence, the control and experimental groups did not differ significantly regarding creative potential across tasks. Finally, an ANCOVA was ran to determine if manipulating the explicit evaluation expectation affected students' stress state. This analysis served as a manipulation check, under the assumption that an evaluative context might increase perceived stress. However, the effect was not statistically significant ($F(1,402)= .644$, $p= .42$), suggesting that the manipulation did not reliably impact participants' stress levels.

Hypothesis 1: Principal effect of explicit evaluation expectation context

For each set of tasks (i.e., tasks completed during the same phase by the participants, *incomplete shapes task/word association task,* vs. *alternative uses task/pictures association*), we ran

MANCOVA analyses using the corresponding creativity scores as DV, while the explicit evaluation expectation context (*expected evaluation* vs. *practice*) was set as the between-subjects factor. Rather than limiting the analysis to Phase 1 results, we compared the Phase 1 results of the *practice context* group with all other results for the same tasks. Specifically, to isolate the effect of expected evaluation on the *Incomplete Shapes task* and the *Word Association task*, we compared the performance of: (1) participants in the *practice* condition who completed these tasks in Phase 1 without any expectation of evaluation (*Order A*), with (2) participants who completed the same tasks in Phase 1 under the *expected evaluation* condition (*Order A*), and all participants who completed these tasks in Phase 2 (*Order B*), which was constantly under evaluative pressure (both for the *practice* and the *expected evaluation* conditions) . Similarly, to examine the effect of expected evaluation on the *Alternative Uses task* and the *Picture Association task*, we compared the performance of: (1) participants in the *practice* condition who completed these tasks in Phase 1 without any expectation of evaluation (*Order B*), with (2) participants who completed the same tasks in Phase 1 under the *expected evaluation* context (*Order B*), and participants who completed these tasks in Phase 2 (*Order A*), which again was constantly under evaluative pressure, regardless of condition.

Phase was included as a covariate to account for potential order effects, as well as age, gender, and digital proficiency level. Results of the MANCOVA analysis, as well as descriptive statistics, are presented in Table 1. Correlation Coefficients between scores are provided in Appendix A for the *expected evaluation* modality and in Appendix B for the *practice* modality[3]. As expected, most dependent variables were significantly intercorrelated within each task and modality, supporting the coherence of the constructs assessed.

Table 1.

[3] Given that tasks were completed under different evaluation contexts depending on the phase, a single correlation matrix for all tasks would not provide meaningful insights. To ensure methodological consistency, separate correlation matrices are presented for task sets completed under comparable evaluation conditions.

Descriptive and inferential statistics for the principal effect of explicit evaluation expectation context on creativity scores from four different tasks

Score	Mean PM (SD)	Mean EEM (SD)	*F*	*Df1*	*Df2*	*p*	η_p^2
IST fluency	.05 (.85)	.05 (.98)	5.052	1	378	.025	.042
IST production time	-.11 (.82)	-.03 (.99)	5.063	1	378	.025	.026
IST originality	-1.18 (.95)	.02 (1.01)	.237	1	378	.627	.024
IST elaboration	-.15 (.84)	-.01 (1.01)	3.136	1	378	.077	.021
IST exploration time	.06 (.88)	-0.03 (.98)	.675	1	378	.412	.032
WA associative thinking	.04 (1.04)	.02 (.95)	4.869	1	378	.028	.207
AUT mean originality	.30 (.83)	-.05 (1.00)	.331	1	347	.567	.063
AUT originality max	.24 (.82)	-.06 (.93)	.457	1	347	.492	.069
AUT fluency	.07 (.76)	.26 (.84)	.561	1	347	.455	.134
PA associative thinking	.10 (.99)	.10 (.96)	.346	1	347	.557	.100

Note. TM = practice modality, EEM = expected evaluation modality, IST = incomplete shapes task, WA = word association, AUT = alternative uses task, PA = pictures association

Results showed that the evaluation context had a principal effect on the incomplete shapes task. Students in the *expected evaluation* modality were significantly less fluent, considering the total response time, than those in the *practice* modality. Further examination of response process data indicated that those participants *expecting evaluation* spent, on average, more time producing their responses. However, the outcomes were not significantly more original or elaborated than in the *practice* modality. No effect of the explicit evaluation expectation context was found on the exploration time for this task. Furthermore, the explicit evaluation expectation context had a principal effect on the global associative thinking score for the word association task. Students in the *expected evaluation* modality made significantly fewer

associations than those in the *practice* modality. Nevertheless, no significant principal effect of explicit evaluation expectation context on any score was found for the AUT, and the pictures association task.

3.2. Hypothesis 2: Moderating effect of school year

For each set of tasks (i.e., tasks completed during the same phase by the participants, *incomplete shapes task/word association task,* vs. *alternative uses task/visual association*), we ran MANCOVA analyses using the corresponding creativity scores as DV, while the explicit evaluation expectation context (*expected evaluation* vs. *practice*), school year and their interaction were included as between-subjects factors. Covariates included age, gender, phase, and digital proficiency level. To enhance interpretability, only significant interactions are illustrated in Figures 1 and 2. Descriptive statistics are presented in Table 2, and all tested effects, including non-significant ones, are fully reported in Table 3.

Table 2.
Descriptive statistics for the moderating effect of school year on the impact of explicit evaluation expectation context on creativity scores from four different tasks

Score	Mean (SD)									
	ISCED 2						ISCED 3			
	1st year		2nd year		3rd year		1st year		2nd year	
	PM	EEM	PM	EEM	PM	EEM	PM	EEM	PM	EEM
IST fluency	-.10 (.59)	.47 (1.08)	-.03 (.73)	.06 (.97)	-.19 (1.15)	-.21 (.81)	.31 (.87)	-.04 (.94)	.29 (.73)	-.01 (.97)
IST production time	-.12 (.79)	-.40 (1.07)	-.04 (.60)	-.09 (1.06)	.19 (.99)	.19 (.88)	-.46 (.91)	.09 (.93)	-.13 (.70)	.05 (.91)
IST originality	-.53 (1.12)	-.39 (.92)	-.20 (.90)	-.17 (.98)	.30 (.87)	.24 (.93)	-.31 (.91)	.17 (1.02)	-.32 (.73)	.24 (1.05)
IST elaboration	-.36 (.90)	-.46 (.98)	.01 (.75)	-.12 (.75)	.16 (.91)	.18 (.90)	-.29 (.90)	.25 (.95)	-.30 (.96)	.13 (1.02)
IST exploration time	-.42 (.79)	-.54 (1.06)	.06 (.75)	-.05 (.83)	.12 (1.01)	.35 (.86)	.17 (.98)	.08 (1.00)	.37 (.66)	-.03 (.95)
WA associative thinking	-.27 (1.00)	-.45 (.88)	.12 (.98)	-.18 (.86)	.15 (.70)	-.04 (.96)	-.15 (1.11)	.31 (1.01)	.31 (1.32)	.51 (.78)
AUT mean originality	.17 (.85)	-.04 (.99)	.20 (.66)	.11 (1.27)	.55 (1.02)	-.58 (.83)	.37 (.72)	.07 (.84)	.18 (.79)	.17 (.89)

AUT originality max	.07	-.31	-.06	.10	.53	-.38	.52	.10	.15	.14
	(.96)	(.81)	(.85)	(1.15)	(.79)	(.89)	(.78)	(.75)	(.47)	(.89)
AUT fluency	-.21	-.31	-.38	.10	.17	.42	.26	.54	.58	.44
	(.67)	(.64)	(.45)	(.82)	(.86)	(.82)	(.40)	(.83)	(.89)	(.79)
PA associative thinking	-58	-.39	.10	.03	.50	.05	.33	.33	.33	.46
	(.97)	(.99)	(1.22)	(.97)	(.80)	(.87)	(.64)	(.92)	(.89)	(.87)

Note. TM = practice modality, EEM = expected evaluation modality, IST = incomplete shapes task, WA = word association, AUT = alternative uses task, PA = pictures association

Table 3.
Inferential statistics for the moderating effect of school year on the effect of explicit evaluation expectation context on creativity scores from four different tasks

Score	*F*	*Df1*	*Df2*	*p*	η_p^2
IST fluency	2.192	4	370	.033	.111
IST production time	1.362	4	370	.176	.085
IST originality	1.925	4	370	.085	.087
IST elaboration	1.225	4	370	.247	.086
IST exploration time	.703	4	370	.590	.116
WA associative thinking	1.190	4	370	.315	.139
AUT mean originality	2.164	4	339	.042	.125
AUT originality max	1.578	4	339	.084	.114
AUT fluency	.827	4	339	.217	.192
PA associative thinking	1.355	4	339	.155	.164

Note. TM = practice modality, EEM = expected evaluation modality, IST = incomplete shapes task, WA = word association, AUT = alternative uses task, PA = pictures association
F-values reflect the interaction effect between explicit evaluation context and school year.

Results showed that the effect of explicit evaluation expectation context on fluency for the incomplete shapes task was moderated by school year (Figure 1), whereby the detrimental effect of evaluation expectation (i.e., explicit evaluation expectation context) on fluency increased across school years over and beyond the effect of age (i.e., controlling for age). Nevertheless, no significant interaction has been found regarding this task's exploration time, production time, originality, and elaboration.

The school year also moderated the effect of the explicit evaluation expectation context on mean originality for the AUT (Figure 2). This effect is particularly pronounced in the third

year, where mean originality reaches its highest point in the practice modality while dropping to its lowest level in the expected evaluation modality.

Figure 1

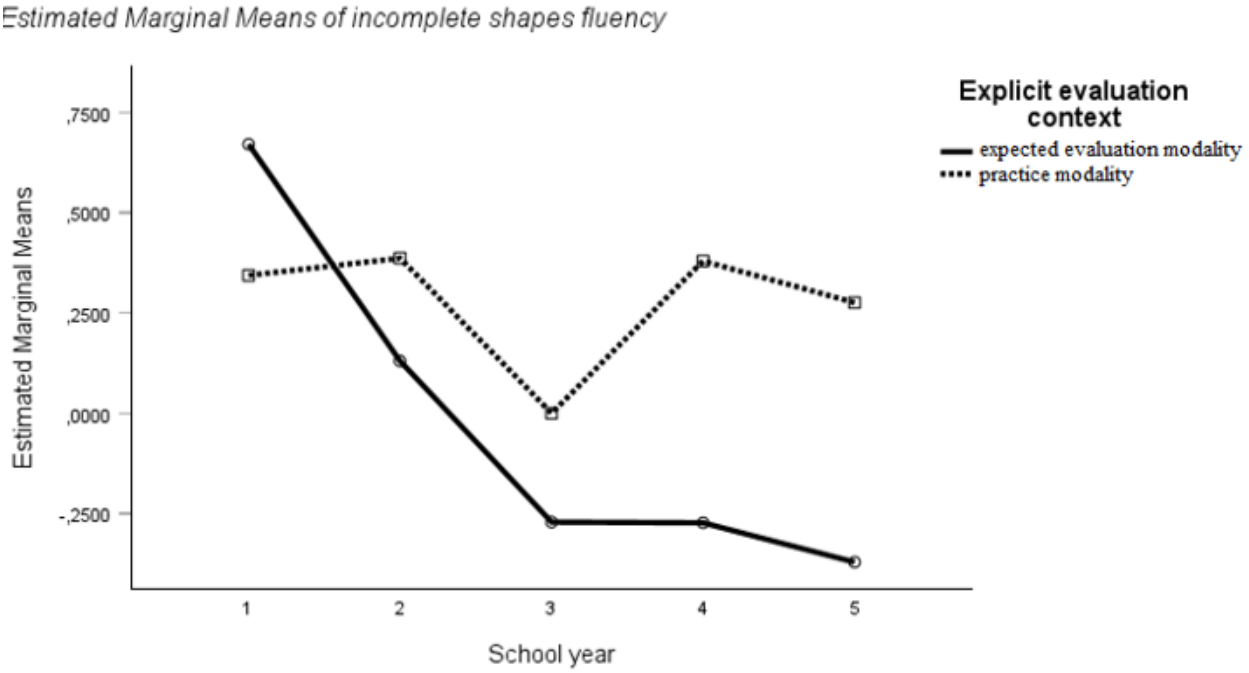
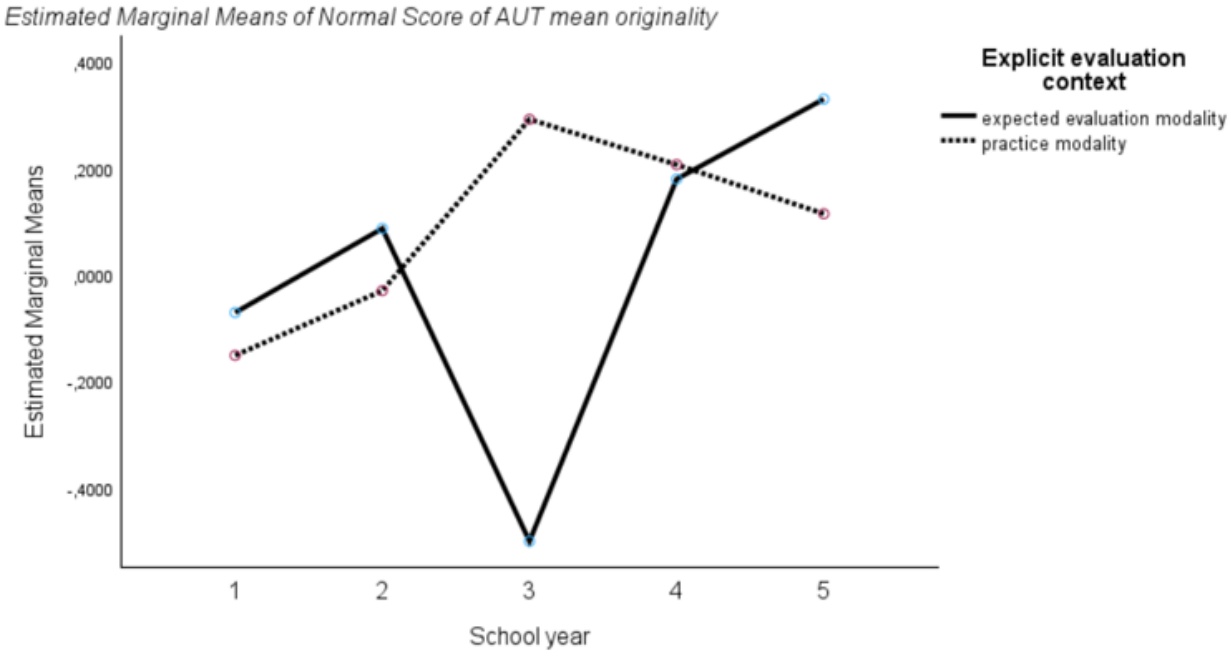


Figure 2



However, no significant interaction between explicit evaluation expectation context and school year on any score was found for the word association and the pictures association task.

3.3. Hypothesis 3: Moderating effect of stress state

For each set of tasks (i.e., tasks completed during the same phase by the participants, *incomplete shapes task/word association task,* vs. *alternative uses task/pictures association*), we ran MANCOVA analyses using the corresponding creativity scores as DV, while the explicit evaluation expectation context (*expected evaluation* vs. *practice*), stress state and their interaction were included as between-subjects factors. Stress state was previously categorized into a group factor by dividing participants into quartiles based on their stress levels, in order to apply an Extreme Groups Approach (EGA), which involves selecting participants from the highest and lowest ends of a continuous distribution (i.e., top and bottom quartiles; Preacher et al., 2005). From a theoretical perspective, the use of EGA is supported by evidence that stress has nonlinear effects on cognitive performance, with both low and high stress levels leading to deviations from the functioning typically observed at moderate levels (Oshri et al., 2022). Moreover, from a methodological standpoint, selecting extreme groups increases statistical power when testing for interaction effects or moderation (Preacher et al., 2005). Covariates included age, gender, phase, and digital proficiency level. Results of the MANCOVA analysis are presented in Table 4, as well as descriptive statistics.

Table 4.
Descriptive and inferential statistics for the moderating effect of stress-state on the effect of explicit evaluation expectation context on creativity scores from four different tasks

Score	Mean down (SD)		Mean top (SD)		*F*	*Df1*	*Df2*	*p*	η_p^2
	PM	EEM	PM	EEM					
IST fluency	.16 (.90)	.08 (.97)	.04 (.86)	.08 (.91)	.008	1	195	.931	.027
IST production time	-.20 (.81)	.003 (1.00)	-.04 (.81)	-.08 (.90)	.267	1	195	.606	.045
IST originality	-.02 (.90)	.14 (1.01)	-.39 (.93)	.05 (.81)	.030	1	195	.863	.031
IST elaboration	-.13 (.82)	.05 (1.02)	-.07 (.82)	-.02 (.95)	.016	1	195	.900	.058
IST exploration time	-.05 (.90)	-.11 (.97)	.15 (.86)	.10 (.93)	1.145	1	195	.286	.056
WA associative thinking	.36 (.85)	-.02 (.90)	-.20 (1.15)	.14 (.93)	5.214	1	195	.023	.116
AUT mean originality	.30 (.83)	-.20 (1.01)	.52 (.91)	.21 (.98)	.030	1	181	.863	.083
AUT Originality max	.19 (.75)	-.13 (.99)	.38 (.78)	.05 (.98)	.135	1	181	.714	.082
AUT fluency	.07 (.82)	.35 (.83)	-.05 (.53)	.31 (.86)	.174	1	181	.677	.118

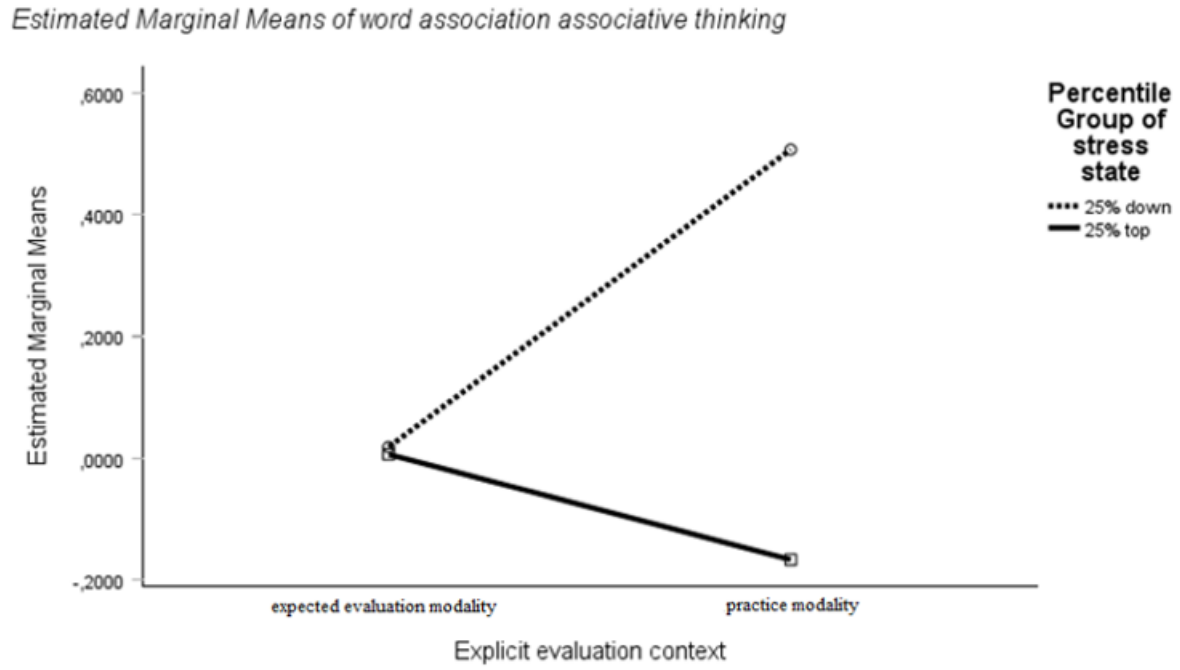
PA associative thinking	.14 (.88)	.07 (1.05)	.57 (.80)	.30 (1.02)	1.013	1	181	.316	.167

Note. TM = practice modality, EEM = expected evaluation modality, IST = incomplete shapes task, WA = word association, AUT = alternative uses task, PA = pictures association

F-values reflect the interaction effect between explicit evaluation context and stress-state.

An interaction was found between the explicit evaluation expectation context and stress state in the word association task score (Figure 2). In the *practice* modality, the most stressed students generated significantly fewer associations than the least stressed ones. There was no sizeable difference in performance in the *expected evaluation* modality according to the stress level.

Figure 3



Nevertheless, no significant interaction between explicit evaluation expectation context and stress state on any score was found for the AUT, the pictures association task, and the incomplete shapes task.

4. Discussion

This study examined how school evaluation context and evaluation stakes influence adolescents' creative potential in four creative tasks targeting associative thinking and creative ideation across visual/graphic and verbal/written modalities. Given the growing international emphasis on creativity in education (Plucker et al., 2015), a better understanding of the effect of

the school evaluation context on creative potential is needed, as it could directly affect educational practices aimed at fostering creativity.

Before proceeding further, it is important to note that intercorrelations between creativity measures derived from different tasks within the same set (i.e., *Incomplete Shapes/Word Association* vs. *Alternative Uses/Pictures Association*) were generally non-significant or very weak. In contrast, strong correlations were mostly observed between subcomponents of the same task (i.e., Incomplete Shapes and Alternative Uses Tasks). These findings suggest that the tasks tap into distinct facets of creativity (i.e., divergent vs associative thinking), which supports the decision to analyze and interpret them separately.

This study demonstrated a principal effect of the explicit evaluation expectation context on three creativity scores. The production time and the fluency score for the incomplete shapes task were impacted. Participants *expecting evaluation* spent more time producing their responses, although the outcomes were not significantly more original or elaborated than in the *practice* modality. In other words, when students anticipated external evaluation, they needed more effort-time to achieve the same level of creative performance as in the practice context, suggesting that evaluation expectation results in a higher "originality cost" (i.e., additional effort-time required to produce an idea of incremental originality, Barbot, 2018). The associative thinking score for the word association task was also impacted: participants *expecting evaluation* made significantly fewer associations than those in the *practice* modality. These findings align with prior research indicating that the school environment, and especially the evaluation expectation, can negatively impact creative performance (e.g., Amabile, 1979; Yuan & Zhou, 2008). This study is nevertheless the first to be conducted in a school context and not limited to divergent thinking, as it also assessed associative thinking. These results suggest that, in a real-world school environment, when students anticipate evaluation, they may inhibit spontaneous exploration and adopt a more cautious, controlled approach, even when the task allows for open-ended responses.

Results also showed a moderating effect of school year on the relation between the explicit evaluation expectation context and the fluency score of the incomplete shapes task. In

the *expected evaluation* modality, students were less fluent as school years increased. Indeed, students from different school years seem not to be affected similarly by the evaluation context, with students in higher years possibly integrating evaluation expectations more strongly than students in lower years. These findings may reflect students' growing internalization of evaluative expectations as they progress through school, leading them to generate a narrower range of ideas during the assessment. A system-level interpretation in terms of "*level of evaluation stakes"* may also be found in a specificity of the Belgian education system: the first two school years are 'low-stakes', in that the daily evaluation is only formative and has little consequences on students' progression. In contrast, the evaluation becomes summative in the subsequent school years, with 'high-stakes' consequences on an end-of-year pass or fail decision. So, beyond developmental effects already well documented (Said-Metwaly et al., 2021) and controlled for in the analyses presented here (i.e., age as a covariate), the 'explicit evaluation expectation context' seems indeed more detrimental to creative potential, particularly to ideational fluency, in a 'high-stakes' context (summative evaluation) than in a 'low-stakes' one (formative evaluation).

Furthermore, a significant interaction between school year and mean originality was found for the AUT. During the first two years, mean originality increases both in practice and expected evaluation modalities, with slightly higher scores in the expected evaluation modality. However, in the third grade, mean originality peaks in the practice modality while it bottoms out in the expected evaluation modality. Then, in the fourth and fifth grades, mean originality decreases in the practice modality while increasing in the expected evaluation modality. These findings suggest that the expectation of evaluation specifically hinders creativity in third grade, coinciding with the shift to summative assessment in the Belgian education system. A plausible explanation is that students become less inclined to take risks with original ideas when anticipating evaluation. This effect appears to diminish over time as they adapt to high-stakes contexts. It is also worth noting that this result aligns with Torrance's (1968) hypothesis regarding the seventh-grade slump, which he associated with an increased fear of others' judgment. The evaluation expectation context may exacerbate this fear, particularly around age

15 (i.e., the average age of third-grade students in Belgium), which corresponds to the age at which this slump is observed in France (Besançon & Lubart, 2015), a cultural context closely related to Belgian Wallonia.

Finally, we found a significant interaction between the explicit evaluation expectation context and stress state on associative thinking (as measured by the word association task). The most stressed students generated more responses in the case of *expected evaluation*, whereas the least stressed generated fewer responses. These results are consistent with Eisenberg and Thompson (2011), who found a beneficial effect of competition, generating stress, on musical creative performance. However, they do not align with previous findings, which describe a detrimental effect of stress on cognitive functioning like attention abilities (Ferreri et al., 2004) and memory retrieval (Wolf, 2017). A reasonable explanation is that individuals expecting evaluation with a higher stress state may be more committed to performing well than those with a low stress state, hence mitigating the detrimental effect of evaluation expectation.

It is worth noting that our findings reveal differential effects of evaluation expectation and stress on divergent thinking and associative thinking. These results underscore the conceptual distinction between these two constructs and align with the cognitive processes underlying them. Associative thinking is typically described as a spontaneous and automatic process that facilitates idea generation through the activation of loosely connected memory networks (Benedek et al., 2012; Wang et al., 2018). In contrast, divergent thinking is considered a more deliberate and controlled process that requires active engagement and exploration of multiple solutions, relying heavily on executive functions and cognitive control (Benedek et al., 2014; Frith et al., 2021). Given this, it is not surprising that evaluation expectation, which involves conscious contextual regulation often leading to reduced risk-taking and more conventional thinking, has a stronger impact on divergent thinking. Conversely, stress is an automatic and largely uncontrollable response to perceived environmental demands, which activates a cascade of physiological and cognitive changes (Kogler et al., 2015). Therefore, it

seems logical that stress levels are more closely related to performance in tasks involving associative thinking, which depend on automatic and less consciously regulated processes.

Despite some strengths of this study, findings can not be fully conclusive regarding the impact of expected evaluation on creative potential. First, the experimental manipulation has shown effects on three out of four creativity tasks (incomplete shapes, alternative uses and word association) and no effects on stress state levels. Interpreting this mitigated pattern of results is complex, and several explanations can be suggested, starting with the distinct nature of the tasks. Indeed, these tasks do not measure the same aspect of creative potential (associative thinking for the pictures association and the word association tasks, and divergent thinking for the others) and use different modalities (graphic/visual or verbal). It can be hypothesized that tasks are not impacted similarly by the evaluation context because of their unique specificities, which is consistent with the findings of Reiter-Palmon et al. (2009) about task and domain specificity of creativity.

Another possible explanation lies in a potential limitation of the effect of the manipulation procedure on the participants' mindset. Indeed, results did not show significant differences in stress state depending on the evaluation context modality: on average, students who took the test in the *expected evaluation* modality did not feel more stressed than those who took the test in the *practice* modality. These unexpected results contradict previous research about the relationship between expected evaluation and stress state (Moyne et al., 2018; Prokofieva et al., 2017). Nevertheless, results showed an effect of explicit evaluation expectation context on creative potential, maybe due to group differences in task engagement. In sum, it is possible that the participants in the explicit evaluation expectation context did not believe the experimenter when they stated that test results could be used during the class councils. Students are often well aware that they cannot be evaluated on skills they did not work in class. Furthermore, the administrated tasks did not resemble typical school tasks, and digital tablets are

not generally used in classrooms for evaluations. Our procedure was also limited by ethical considerations, which involved informing participants that these activities occurred in a scientific research context, perhaps conflicting with the idea of a real school evaluation.

However, we believe solutions exist to solve these issues in further studies. For example, to make the situation less artificial in a school setting, we could use more academic tasks, like those proposed by Gontijo and Fleith (2014). Indeed, they explain that creativity holds an important place in mathematics and devised DT tasks involving finding as many ways as possible to reach a given result. The level of difficulty of such tasks could be adapted to each school year.

To conclude, H1 (principal effect of explicit evaluation expectation context) was partially supported, with significant main effects of evaluation expectation on two out of four tasks. H2 (moderating effect of school year) received partial support, as school year moderated the effect of evaluation on two different scores. H3 (moderating effect of stress-state) was supported only in one task, where stress moderated the effect of evaluation expectation. Despite some possible improvements, this study provides a solid ground and encouraging preliminary findings to further study the role of the school evaluation context on creative thinking. First, its ecological validity was warranted by conducting it within an authentic school setting, allowing for a more accurate understanding of how real-world evaluation context affects creative potential. It also provided new insights into how evaluation contexts can differentially affect creative potential across multiple aspects of creative thinking. Notably, we found that the evaluation expectation increases the time and effort required to reach similar levels of originality in a graphic divergent thinking task, compared to a situation in which evaluation was not expected. It also hinders overall verbal associative thinking skills. Moreover, the moderating effect of school year on ideational fluency suggests that students in later school years may be more susceptible to the constraining effects of evaluation. Along the same lines, the moderating effect of school year on

the mean originality score for the AUT suggests that students in their third year of secondary school are particularly affected by the explicit school evaluation context. This may be explained by the level of evaluation stakes, characterized by a shift from low- to high-stakes evaluations. Interestingly, this study also highlighted that state stress does not uniformly affect creative potential under expected evaluation. Students with higher state stress levels were able to generate more responses in a verbal associative thinking task when expecting evaluation, suggesting that, in the context of evaluation expectation, state stress may motivate people to increase their creative efforts. Together, this study suggests important practical implications for educators in fostering creativity in school settings. Finally, it is also important to remind that our study focused only on two cognitive components of creative potential (i.e., divergent and associative thinking) while other facets of creativity remain unexplored and should be addressed in future research.

5. Conclusion

This study showed that the explicit evaluation expectation context in school had a detrimental effect on several aspects of creative potential, with variations depending on the task and modalities (e.g., verbal vs graphic) considered. However, this effect was not systematic across all measures of creative potential, suggesting that the impact of evaluation expectations may be domain- or modality-specific. Additionally, school level (and indirectly, evaluation stakes) and stress state level mitigated some of these effects. Together, these findings can initiate a much-needed line of work on the evaluative context in school settings and its impact on creative potential. This knowledge base is needed to monitor creativity change and development across school years properly. It can further fuel education debates on optimal conditions for creativity assessment, considering the targeted aspect of creativity to measure and its interaction with the explicit evaluative context and the level of evaluation stakes

References

Acar, S., Burnett, C., & Cabra, J. F. (2017). Ingredients of Creativity: Originality and More. *Creativity Research Journal*, *29*(2), 133–144. https://doi.org/10.1080/10400419.2017.1302776

Acar, S., Organisciak, P., & Dumas, D. (2024). Automated Scoring of Figural Tests of Creativity with Computer Vision. *The Journal of Creative Behavior*. https://doi.org/10.1002/jocb.677

Amabile, T. (2011). *Componential theory of creativity*. Harvard Business School Boston, MA.

Amabile, T. M. (1979). *Effects of external evaluation on artistic creativity*. *37*(2), 221–233. https://doi.org/10.1037/0022-3514.37.2.221

Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, *43*(5), 997–1013. https://doi.org/10.1037/0022-3514.43.5.997

Amabile, T. M., Goldfarb, P., & Brackfleld, S. C. (1990). Social influences on creativity: Evaluation, coaction, and surveillance. *Creativity Research Journal*, *3*(1), 6–21. https://doi.org/10.1080/10400419009534330

Amabile, T. M., & Pratt, M. G. (2016). The dynamic componential model of creativity and innovation in organizations: Making progress, making meaning. *Research in Organizational Behavior*, *36*, 157–183. https://doi.org/10.1016/j.riob.2016.10.001

Archer, E. (2017). The Assessment Purpose Triangle: Balancing the Purposes of Educational Assessment. *Frontiers in Education*, *2*. https://doi.org/10.3389/feduc.2017.00041

Baer, J. (1997). Gender Differences in the Effects of Anticipated Evaluation on Creativity. *Creativity Research Journal*, *10*(1), 25–31. https://doi.org/10.1207/s15326934crj1001_3

Barbot, B. (2018a). "Generic" Creativity as a Predictor or Outcome of Identity Development? *Creativity Theories–Research–Applications*, *2*(5), 159–164. https://doi.org/10.1515/ctra-2018-0013

Barbot, B. (2018b). The Dynamics of Creative Ideation: Introducing a New Assessment Paradigm. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2018.02529

Barbot, B. (2019). Measuring creativity change and development. *Psychology of Aesthetics, Creativity, and the Arts.*, *13*(2), 203–210. https://doi.org/10.1037/aca0000232

Barbot, B. (2020). Creativity and Self-esteem in Adolescence: A Study of Their Domain-Specific, Multivariate Relationships. *The Journal of Creative Behavior*, *54*(2), 279–292. https://doi.org/10.1002/jocb.365

Barbot, B., & Lubart, T. (2012). Adolescence, créativité et transformation de Soi. *Enfance*, *3*(3), 299–312. https://doi.org/10.3917/enf1.123.0299

Barbot, B., Lubart, T. I., & Besançon, M. (2016). "Peaks, Slumps, and Bumps": Individual Differences in the Development of Creativity in Children and Adolescents. *New Directions for Child and Adolescent Development*, *2016*(151), 33–45. https://doi.org/10.1002/cad.20152

Barbot, B., & Rogh, W. (2020). Developmental Trends in Creative Abilities and Potentials. In S. Pritzker & M. Runco (Eds.), *Encyclopedia of Creativity (Third Edition)* (pp. 323–326). Academic Press. https://doi.org/10.1016/B978-0-12-809324-5.06167-8

Barbot, B., & Said-Metwaly, S. (2021). Is There Really a Creativity Crisis? A Critical Review and Meta-analytic Re-Appraisal. *The Journal of Creative Behavior*, *55*(3), 696–709.

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*(3), 211–227. https://doi.org/10.3758/BF03196968

Bartis, S., Szymanski, K., & Harkins, S. G. (1988). Evaluation and Performance: A Two-Edged Knife. *Personality and Social Psychology Bulletin*, *14*(2), 242–251. https://doi.org/10.1177/0146167288142003

Beaty, R. E., & Kenett, Y. N. (2023). Associative thinking at the core of creativity. *Trends in Cognitive Sciences*, *27*(7), 671–683. https://doi.org/10.1016/j.tics.2023.04.004

Beghetto, R. A., & Sriraman, B. (2017). Introduction: Creative Contradictions in Education. In *Creative Contradictions in Education* (pp. xi–xiii). Springer International Publishing.

Benedek, M., Jauk, E., Sommer, M., Arendasy, M., & Neubauer, A. C. (2014). Intelligence, creativity, and cognitive control: The common and differential involvement of executive functions in intelligence and creativity. *Intelligence*, *46*. https://doi.org/10.1016/j.intell.2014.05.007

Benedek, M., Könen, T., & Neubauer, A. C. (2012). Associative abilities underlying creativity. *Psychology of Aesthetics, Creativity, and the Arts*, *6*(3), 273–281. https://doi.org/10.1037/a0027059

Benedek, M., Mühlmann, C., Jauk, E., & Neubauer, A. C. (2013). Assessment of divergent thinking by means of the subjective top-scoring method: Effects of the number of top-ideas and time-on-task on reliability and validity. *Psychology of Aesthetics, Creativity, and the Arts*, *7*(4), 341–349. https://doi.org/10.1037/a0033644

Besançon, M., & Lubart, T. (2015). *La créativité de l'enfant: Évaluation et développement*. Mardaga.

Bonnardel, N. (2009). Activités de conception et créativité: De l'analyse des facteurs cognitifs à l'assistance aux activités de conception créatives. *Le travail humain*, *72*(1), 5–22. https://doi.org/10.3917/th.721.0005

Borger, L., Eklöf, H., Johansson, S., & Strietholt, R. (2025). The issue of test-taking motivation in low- and high-stakes tests: Are students underachieving in PISA? *Learning and Individual Differences*, *122*, 102722. https://doi.org/10.1016/j.lindif.2025.102722

Byron, K., & Khazanchi, S. (2011). A Meta-Analytic Investigation of the Relationship of State and Trait Anxiety to Performance on Figural and Verbal Creative Tasks. *Personality and Social Psychology Bulletin*, *37*(2), 269–283. https://doi.org/10.1177/0146167210392788

Camarda, A., Bouhours, L., Osmont, A., Le Masson, P., Weil, B., Borst, G., & Cassotti, M. (2021). Opposite Effect of Social Evaluation on Creative Idea Generation in Early and Middle Adolescents. *Creativity Research Journal*, *33*(4), 399–410. https://doi.org/10.1080/10400419.2021.1902174

Cebeci, S. M., & Acar, S. (2025). Development and Validation of the Cebeci Test of Creativity: A Computerized Test of Figural Creativity. *The Journal of Creative Behavior*, *59*(2), e70030. https://doi.org/10.1002/jocb.70030

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences, 3rd ed* (pp. xxviii, 703). Lawrence Erlbaum Associates Publishers.

de Jesus, S. N., Rus, C. L., Lens, W., & Imaginário, S. (2013). Intrinsic Motivation and Creativity Related to Product: A Meta-analysis of the Studies Published Between 1990–2010. *Creativity Research Journal*, *25*(1), 80–84. https://doi.org/10.1080/10400419.2013.752235

Dumas, D., Organisciak, P., & Doherty, M. (2021). Measuring divergent thinking originality with human raters and text-mining models: A psychometric comparison of methods. *Psychology of Aesthetics, Creativity, and the Arts*, *15*(4), 645–663. https://doi.org/10.1037/aca0000319

Duval, S., & Wicklund, R. A. (1972). *A Theory of Objective Self Awareness*. Academic Press.

Eisenberg, J., & Thompson, W. F. (2011). The Effects of Competition on Improvisers' Motivation, Stress, and Creative Performance. *Creativity Research Journal*, *23*(2), 129–136. https://doi.org/10.1080/10400419.2011.571185

Eisenberger, R., & Rhoades, L. (2001). Incremental effects of reward on creativity. *Journal of Personality and Social Psychology*, *81*(4), 728–741. https://doi.org/10.1037/0022-3514.81.4.728

Fédération Wallonie-Bruxelles. (2020). *Arrêté du Gouvernement de la Communauté française établissant le classement des implantations de l'enseignement fondamental et de l'enseignement secondaire en application de l'article 4 du décret du 30 avril 2009, organisant un encadrement différencié au sein des établissements scolaires de la Communauté française afin d'assurer à chaque élève des chances égales d'émancipation sociale dans un environnement pédagogique de qualité*. https://www.gallilex.cfwb.be/document/pdf/48085_000.pdf

Fédération Wallonie-Bruxelles. (2023). *Informations relatives à la mise en œuvre du tronc commun durant l'année scolaire 2023-2024*. http://www.enseignement.be/upload/circulaires/000000000004/FWB%20-%20Circulaire%208936%20(9191_20230601_093458).pdf

Fédération Wallonie-Bruxelles. (2024). *Indice socioéconomique—Explication de la procédure de calcul*.

Ferreri, F., Peretti, C., & Mattei, P. (2004). Stress et cognition. *Act. Méd. Int. Psychiatr*, *21*, 9–10.

Forthmann, B., Lips, C., Szardenings, C., Scharfen, J., & Holling, H. (2020). Are Speedy Brains Needed when Divergent Thinking is Speeded—Or Unspeeded? *The Journal of Creative Behavior*, *54*(1), 123–133. https://doi.org/10.1002/jocb.350

Frith, E., Kane, M. J., Welhaf, M. S., Christensen, A. P., Silvia, P. J., & Beaty, R. E. (2021). Keeping Creativity under Control: Contributions of Attention Control and Fluid Intelligence to Divergent Thinking. *Creativity Research Journal*, *33*(2), 138–157. https://doi.org/10.1080/10400419.2020.1855906

Golab, D., & Barbot, B. (2024). Building Narratives Through Empathy: The Role of Empathy Mechanisms and Associative Thinking in Creative Writing. *The Journal of Creative Behavior*, *58*(4), 739–754. https://doi.org/10.1002/jocb.1516

Gontijo, C. H., & Fleith, D. de S. (2014). Assessment of Creativity in Mathematics. In *Theory and Practice of Creativity Measurement*. Routledge.

Gralewski, J., & Karwowski, M. (2018). Are Teachers' Implicit Theories of Creativity Related to the Recognition of Their Students' Creativity? *The Journal of Creative Behavior*, *52*(2), 156–167. https://doi.org/10.1002/jocb.140

Gralewski, J., Lebuda, I., Gajda, A., Jankowska, D. M., & Wiśniewska, E. (2016). Slumps and Jumps: Another Look at Developmental Changes in Creative Abilities. *Creativity. Theories – Research - Applications*, *3*(1), 152–177. https://doi.org/10.1515/ctra-2016-0011

Green, A. E., Beaty, R. E., Kenett, Y. N., & Kaufman, J. C. (2023). The Process Definition of Creativity. *Creativity Research Journal*, *0*(0), 1–29. https://doi.org/10.1080/10400419.2023.2254573

Guilford, J. P. (1950). Creativity. *American Psychologist*, *5*(9), 444–454. https://doi.org/10.1037/h0063487

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

He, W., & Wong, W. (2015). Creativity slump and school transition stress: A sequential study from the perspective of the cognitive-relational theory of stress. *Learning and Individual Differences*, *43*, 185–190. https://doi.org/10.1016/j.lindif.2015.08.034

Hu, W., Wang, X., Yi, L. Y. X., & Runco, M. A. (2018). Creative self-efficacy as moderator of the influence of evaluation on artistic creativity. *The International Journal of Creativity & Problem Solving*, *28*(2), 39–55.

Jirásek, M., & Sudzina, F. (2020). *Big Five Personality Traits and Creativity*. *24*(3). https://www.muni.cz/vyzkum/publikace/1693136

Karwowski, M. (2022). *School does not kill creativity*. *27*(3), 263–275. https://doi.org/10.1027/1016-9040/a000449

Karwowski, M., Czerwonka, M., Wiśniewska, E., & Forthmann, B. (2021). How Is Intelligence Test Performance Associated with Creative Achievement? A Meta-Analysis. *Journal of Intelligence*, *9*(2), Article 2. https://doi.org/10.3390/jintelligence9020028

Karwowski, M., Lebuda, I., & Beghetto, R. A. (2019). Creative self-beliefs. In *The Cambridge handbook of creativity, 2nd ed* (pp. 396–417). Cambridge University Press. https://doi.org/10.1017/9781316979839.021

Kaufman, J. C., Beghetto, R. A., & Dilley, A. (2016). Understanding Creativity in the Schools. In A. A. Lipnevich, F. Preckel, & R. D. Roberts (Eds.), *Psychosocial Skills and School Systems in the 21st Century: Theory, Research, and Practice* (pp. 133–153). Springer International Publishing. https://doi.org/10.1007/978-3-319-28606-8_6

Kogler, L., Müller, V. I., Chang, A., Eickhoff, S. B., Fox, P. T., Gur, R. C., & Derntl, B. (2015). Psychosocial versus physiological stress—Meta-analyses on deactivations and activations of the neural correlates of stress reactions. *NeuroImage*, *119*. https://doi.org/10.1016/j.neuroimage.2015.06.059

Kruglanski, A. W., Friedman, I., & Zeevi, G. (1971). The effects of extrinsic incentive on some qualitative aspects of task performance. *Journal of Personality*, *39*(4), 606–617. https://doi.org/10.1111/j.1467-6494.1971.tb00066.x

Lei, W., Deng, W., Zhu, R., Runco, M. A., Dai, D. Y., & Hu, W. (2021). Does Cognitive Style Moderate Expected Evaluation and Adolescents' Creative Performance: An Empirical Study. *The Journal of Creative Behavior*, *55*(1), 120–129. https://doi.org/10.1002/jocb.439

Lepper, M. R., Greene, D., & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the 'overjustification' hypothesis. *Journal of Personality and Social Psychology*, *28*(1), 129–137. https://doi.org/10.1037/h0035519

Lubart, T., & Besançon, M. (2017). On the Measurement and Mismeasurement of Creativity. In R. A. Beghetto & B. Sriraman (Eds.), *Creative Contradictions in Education: Cross Disciplinary Paradoxes and Perspectives* (pp. 333–348). Springer International Publishing. https://doi.org/10.1007/978-3-319-21924-0_18

Lubart, T., & Guignard, J.-H. (2004). The Generality-Specificity of Creativity: A Multivariate Approach. In *Creativity: From potential to realization* (pp. 43–56). American Psychological Association. https://doi.org/10.1037/10692-004

Miller, A. L. (2007). Creativity and cognitive style: The relationship between field-dependence-independence, expected evaluation, and creative performance. *Psychology of Aesthetics, Creativity, and the Arts*, *1*(4), 243–246. https://doi.org/10.1037/1931-3896.1.4.243

Moyne, A., Barthod-Malat, Y., & Kubiszewski, V. (2018). Stress scolaire des collégiens et des lycéens: Effet de l'appréhension de l'orientation et de la pression de la note. *Pratiques Psychologiques*, *24*(2), 177–194. https://doi.org/10.1016/j.prps.2017.07.001

Nijstad, B. A., De Dreu, C. K. W., Rietzschel, E. F., & Baas, M. (2010). The dual pathway to creativity model: Creative ideation as a function of flexibility and persistence. *European Review of Social Psychology*, *21*(1), 34–77. https://doi.org/10.1080/10463281003765323

Nikkola, T., Kangas, J., & Reunamo, J. (2024). Children's creative participation as a precursor of 21st century skills in Finnish early childhood education and care context. *Learning and Individual Differences*, *111*, 102437. https://doi.org/10.1016/j.lindif.2024.102437

Organisciak, P., Acar, S., Dumas, D., & Berthiaume, K. (2022). *Beyond Semantic Distance: Automated Scoring of Divergent Thinking Greatly Improves with Large Language Models*. https://doi.org/10.13140/RG.2.2.32393.31840

Oshri, A., Cui, Z., Owens, M. M., Carvalho, C. A., & Sweet, L. (2022). Low-to-moderate level of perceived stress strengthens working memory: Testing the hormesis hypothesis through neural activation. *Neuropsychologia*, *176*, 108354. https://doi.org/10.1016/j.neuropsychologia.2022.108354

Pannier, I. (2005). Pour en finir (ou presque) avec les notes—Évaluer par les compétences. *Les Cahiers pédagogiques*, *438*, 15–16.

Patterson, J. D., Barbot, B., Lloyd-Cox, J., & Beaty, R. E. (2024). AuDrA: An automated drawing assessment platform for evaluating creativity. *Behavior Research Methods*, *56*(4), 3619–3636. https://doi.org/10.3758/s13428-023-02258-3

Payne, R. W., & Friedlander, D. (1962). A short battery of simple tests for measuring overinclusive thinking. *The British Journal of Psychiatry*, *108*(454), 362–367.

Plucker, J. A., Kaufman, J. C., & Beghetto, R. A. (2015). *What we know about creativity. Washington, DC: Partnership for 21st Century Skills*.

Plucker, J. A., Qian, M., & Wang, S. (2011). Is Originality in the Eye of the Beholder? Comparison of Scoring Techniques in the Assessment of Divergent Thinking. *The Journal of Creative Behavior*, *45*(1), 1–22. https://doi.org/10.1002/j.2162-6057.2011.tb01081.x

Preacher, K. J., Rucker, D. D., MacCallum, R. C., & Nicewander, W. A. (2005). Use of the Extreme Groups Approach: A Critical Reexamination and New Recommendations. *Psychological Methods*, *10*(2). https://doi.org/10.1037/1082-989x.10.2.178

Prokofieva, V., Brandt-Pomares, P., Velay, J.-L., Hérold, J.-F., & Kostromina, S. (2017). Stress de l'évaluation scolaire: Un nouveau regard sur un problème ancien. *Recherches & éducations*, *18*, Article 18. https://doi.org/10.4000/rechercheseducations.4657

Reiter-Palmon, R., Forthmann, B., & Barbot, B. (2019). Scoring divergent thinking tests: A review and systematic framework. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(2), 144.

Reiter-Palmon, R., Illies, M. Y., Kobe Cross, L., Buboltz, C., & Nimps, T. (2009). Creativity and domain specificity: The effect of task type on multiple indexes of creative problem-solving. *Psychology of Aesthetics, Creativity, and the Arts*, *3*(2), 73–80. https://doi.org/10.1037/a0013410

Rubenstein, L. D., Ridgley, L. M., Callan, G. L., Karami, S., & Ehlinger, J. (2018). How teachers perceive factors that influence creativity development: Applying a Social Cognitive Theory perspective. *Teaching and Teacher Education*, *70*, 100–110. https://doi.org/10.1016/j.tate.2017.11.012

Runco, M. A., & Acar, S. (2012). Divergent Thinking as an Indicator of Creative Potential. *Creativity Research Journal*, *24*(1), 66–75. https://doi.org/10.1080/10400419.2012.652929

Runco, M. A., & Jaeger, G. J. (2012). The Standard Definition of Creativity. *Creativity Research Journal*, *24*(1), 92–96. https://doi.org/10.1080/10400419.2012.650092

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Said-Metwaly, S., Fernández-Castilla, B., Kyndt, E., Van den Noortgate, W., & Barbot, B. (2021). Does the Fourth-Grade Slump in Creativity Actually Exist? A Meta-analysis of the Development of Divergent Thinking in School-Age Children and Adolescents. *Educational Psychology Review*, *33*(1), 275–298. https://doi.org/10.1007/s10648-020-09547-9

Said-Metwaly, S., Taylor, C. L., Camarda, A., & Barbot, B. (2022). Divergent thinking and creative achievement—How strong is the link? An updated meta-analysis. *Psychology of Aesthetics, Creativity, and the Arts*, No Pagination Specified-No Pagination Specified. https://doi.org/10.1037/aca0000507

Saretzki, J., Forthmann, B., & Benedek, M. (2024). A systematic quantitative review of divergent thinking assessments. *Psychology of Aesthetics, Creativity, and the Arts*. https://doi.org/10.1037/aca0000691

Shalley, C. E. (1995). Effects of coaction, expected evaluation, and goal setting on creativity and productivity. *Academy of Management Journal*, *38*(2), 483–503.

Shalley, C. E., & Perry-Smith, J. E. (2001). Effects of Social-Psychological Factors on Creative Performance: The Role of Informational and Controlling Expected Evaluation and Modeling Experience. *Organizational Behavior and Human Decision Processes*, *84*(1), 1–22. https://doi.org/10.1006/obhd.2000.2918

Silvia, P. J., Beaty, R. E., & Nusbaum, E. C. (2013). Verbal fluency and creativity: General and specific contributions of broad retrieval ability (Gr) factors to divergent thinking. *Intelligence*, *41*(5), 328–340. https://doi.org/10.1016/j.intell.2013.05.004

Silvia, P. J., & Phillips, A. G. (2004). Self-Awareness, Self-Evaluation, and Creativity. *Personality and Social Psychology Bulletin*, *30*(8), 1009–1017. https://doi.org/10.1177/0146167204264073

Stables, K. (2009). Educating for environmental sustainability and educating for creativity: Actively compatible or missed opportunities? *International Journal of Technology and Design Education*, *19*(2), 199–219. https://doi.org/10.1007/s10798-008-9077-1

Sternberg, R. J., & Lubart, T. I. (1995). *Defying the crowd: Cultivating creativity in a culture of conformity* (pp. ix, 326). Free Press.

Sternberg, R. J., & Lubart, T. I. (2004). Creative giftedness: A multivariate investment approach. *Creativity and Giftedness*, *10*, 141.

Taylor, C. L., & Barbot, B. (2021). *Dual pathways in creative writing processes*. *18*(2), 196–208. https://doi.org/10.1037/aca0000415

Taylor, C. L., Said-Metwaly, S., Camarda, A., & Barbot, B. (2024). Gender differences and variability in creative ability: A systematic review and meta-analysis of the greater male variability hypothesis in creativity. *Journal of Personality and Social Psychology*, *126*(6), 1161–1179. https://doi.org/10.1037/pspp0000484

Torrance, E. P. (1968). A Longitudinal Examination of the Fourth Grade Slump in Creativity. *Gifted Child Quarterly*, *12*(4), 195–199. https://doi.org/10.1177/001698626801200401

Torrance, E. P. (2008). *Torrance Tests of Creative Thinking: Norms-Technical Manual, Verbal Forms A and B* (Scholastic Testing Service).

Walia, C. (2019). A Dynamic Definition of Creativity. *Creativity Research Journal*, *31*(3), 237–247. https://doi.org/10.1080/10400419.2019.1641787

Wang, P., Wijnants, M. L., & Ritter, S. M. (2018). What Enables Novel Thoughts? The Temporal Structure of Associations and Its Relationship to Divergent Thinking. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01771

Wolf, O. T. (2017). Stress and memory retrieval: Mechanisms and consequences. *Current Opinion in Behavioral Sciences*, *14*, 40–46. https://doi.org/10.1016/j.cobeha.2016.12.001

Xue, Y., Gu, C., Wu, J., Dai, D. Y., Mu, X., & Zhou, Z. (2020). The Effects of Extrinsic Motivation on Scientific and Artistic Creativity among Middle School Students. *The Journal of Creative Behavior*, *54*(1), 37–50. https://doi.org/10.1002/jocb.239

Yi, X., Hu, W., Plucker, J. A., & McWilliams, J. (2013). Is There a Developmental Slump in Creativity in China? The Relationship Between Organizational Climate and Creativity Development in Chinese Adolescents. *The Journal of Creative Behavior*, *47*(1), 22–40. https://doi.org/10.1002/jocb.21

Yuan, F., & Zhou, J. (2008). Differential Effects of Expected External Evaluation on Different Parts of the Creative Idea Production Process and on Final Product Creativity. *Creativity Research Journal*, *20*(4), 391–403. https://doi.org/10.1080/10400410802391827

Appendix A

Correlation coefficients for the incomplete shapes and the word association tasks

Table 5

Correlation coefficients for the incomplete shape and the word association tasks in the expected evaluation modality

Variable	1	2	3	4	5
1. IST production time					
2. IST elaboration	.906**				
3. IST fluency	-.917**	-.803**			
4. IST originality	.504**	.606**	-.489**		
5. IST exploration time	.374**	.241**	-.483**	.204**	
6. WA associative thinking	-.053	.037	.094	.133*	-.095

** p < .01 (2-tailed), * p < .05 (2-tailed), IST = incomplete shapes task, WA = word association

Table 6

Correlation coefficients for the incomplete shape and the word association tasks in the practice modality

Variable	1	2	3	4	5
1. IST production time					
2. IST elaboration	.883**				
3. IST fluency	-.903**	-.779**			
4. IST originality	.481**	.552**	-.487**		
5. IST exploration time	.423**	.341**	-.464**	.195*	
6. WA associative thinking	.046	.142	.004	.056	.034

** p < .01 (2-tailed), * p < .05 (2-tailed), IST = incomplete shapes task, WA = word association

Appendix B

Correlation coefficients for the AUT and the pictures association task

Table 7

Correlation coefficients for the AUT and the pictures association task in the practice modality

Variable	1	2	3
1. AUT mean originality			
2. AUT originality max	.747**		
3. AUT fluency	-,200	.189	
4. PAT associative thinking	,097	.123	.272**

** p < .01 (2-tailed), AUT = alternative uses task, PAT = pictures association task

Table 8

Correlation coefficients for the AUT and the pictures association task in the expected evaluation modality

Variable	1	2	3
1. AUT mean originality			
2. AUT originality max	.831**		
3. AUT fluency	-.126*	.182**	
4. PAT associative thinking	,148*	.206**	.259**

** p < .01 (2-tailed), * p < .05 (2-tailed), AUT = alternative uses task, PAT = pictures association task