Gender and creative ability: Mean differences and performance variability

Zorana Ivcevic[1], Olena Zyga[2], Jessica D. Hoffmann[1], and Raquel Palomera[3]

[1] Yale Center for Emotional Intelligence, Yale University

[2] Department of Psychiatry and Behavioral Sciences, Stanford University

[3] Department of Education, University of Cantabria

Author Note

Correspondence concerning this article should be addressed to Zorana Ivcevic, Yale Center for Emotional Intelligence, Yale University, 350 George St., New Haven, CT 06511.

Zorana.ivcevic@yale.edu

Abstract

Studies examining gender differences in creative abilities remain inconsistent and call for further research using diverse measures. The current study examined gender differences in mean performance, as well as differences in variability on different performance criteria for male and female participants. The study includes a sample of children and adolescents (N = 678; 1st-10th grade in primary and secondary schools) and uses tasks in verbal (storytelling) and figural domains (incomplete drawings) scored for six performance criteria (creativity, novelty/originality, likability, elaboration, humor, and emotion use). Similar to previous research, when mean differences in performance emerged, they favored girls; however, these differences tended to be small and not consistent across all performance criteria. The greater male variability hypothesis in performance was not supported. These results point to the limits of the greater male variability hypothesis and the need to examine gender differences in creative abilities across a broad range of tasks and domains and investigate developmental factors that could be related to the emergence of gender differences.

Keywords: creative thinking, verbal creativity domain, figural creativity domain, gender differences, greater male variability hypothesis

Gender and creative ability: Mean differences and variability

Gender is an important factor in the expressed trajectory of creative development (Karwowski et al., 2016; Lau & Cheung, 2010). However, results about the relationship between gender and creative ability remain inconsistent and call for further research using different measures. Overall, previous results on mean differences in performance on tests of creative ability led creativity scholars to a conclusion about a "relative equality" between males and females (Baer & Kaufman, 2008; Kogan, 1974; Urban, 1991). Some studies find a significant difference and when gender differences in mean creative ability are observed, females usually perform better (Baer & Kaufman, 2008 for a review).

A more recent line of research examines not only gender differences based on mean scores but also variability on performance criteria for male and female participants, inspired by research on other intellectual abilities and personality. Methodologically, two main approaches to analyzing variability include calculating male/female variance ratios (VRs) and examining proportions of male and female participants in different regions of the score distribution (comparing male/female ratio in lower, middle, and upper portions of the score distribution). Previous work has shown that gender ratios in the tail ends of the distribution are critical in understanding gender differences in intellectual abilities (Hedges & Friedman, 1993). Greater male variability, both in terms of VRs and greater proportion of males at the low and high end of score distributions, was found across various areas of functioning, such as intellectual abilities (Johnson et al., 2008; Feingold, 1992; Hedges & Nowell, 1995) or self-esteem (Kling et al., 1999).

In relation to creative abilities, a number of studies found support for the greater male variability hypothesis (He, 2018; He & Wong, 2011; Karwowski et al., 2016). These initial

studies measured creative abilities in terms of performance on the Test of Creative Thinking – Drawing Production (TCT-DP), a figural test taking a holistic approach to assessment of creative ability (Urban & Jellen, 1996). For instance, in a longitudinal study, He (2018) found support for higher male variability across scoring criteria for the TCT-DP. However, when different measures are used, such as storytelling and divergent thinking tasks, evidence for greater male variability is not found (Taylor & Barbot, 2021).

Given these previous findings and that past work has mostly focused on administering figural creativity tasks, with limited work assessing creative ability across task domains and multiple performance criteria, the present study addresses gender differences across figural and verbal domains and in relation to different performance criteria, including creativity, novelty/originality, likeability, elaboration, humor, and emotion use. The current study examines gender differences both in terms of mean scores and variability in a sample of primary and secondary school children.

Mean Gender Differences in Creative Abilities

Although we have learned much about creative thinking and potential – from the role of emotion states in creative thinking (Baas et al., 2008) to the relationship between creative abilities and general intellectual ability (Kim, 2005) to the role of mindfulness (Lebuda et al., 2016) —research on gender differences in creative thinking comparatively has not received as much attention (Baer & Kaufman, 2008). This may be in part due to the fact that nearly 50 years ago Kogan (1974) had already noted “relative equality” in creative thinking among males and females in an extensive review of existing research. Subsequent results continued to support the notion of relative equality when mean differences in performance on creative thinking tests are evaluated. Consistent findings suggest similarity in performance across samples and different

ages pointing to a relative equality in creative ability in boys/men and girls/women (Bender et al., 2013; Cheung & Lau, 2010; Hong & Milgram, 2010; Karwowski et al., 2013; Kaufman et al., 2010; Sayed & Mohamed, 2013; Stoltzfus et al., 2011; Taylor & Barbot, 2021).

In the last two decades, there have been three major reviews on gender differences in creative thinking (Baer & Kaufman, 2008; Pagnani, 2011; Runco et al., 2010). The main takeaway from these reviews is that about half of the studies conducted showed no significant differences between males and females on creative ability. The other half of the studies showed mixed results, potentially suggesting higher creative thinking ability in females. The suggestion of higher performance by female participants parallels findings in personality research showing that females score higher on the Openness to Experience trait, which can be described as the personality disposition toward creativity (Costa et al., 2001; Karwowski et al., 2013; Misra, 2003).

The findings of relative equality are consistent for different performance criteria on divergent thinking tests (fluency, originality, flexibility). Similar performance across gender groups is also found when creative output on laboratory tasks is assessed by outside observers. When studies have asked participants to create something (i.e., poem, story, drawing) and subsequently rated for creativity by blinded experts using the Consensual Assessment Technique, no significant gender differences in mean performance are found (Amabile, 1982, 1983; Baer, 1993; Kaufman et al., 2004; Kaufman et al., 2010; Taylor & Barbot, 2021).

One limitation of existing research is in the measures used to assess creative thinking. In their review, Baer and Kaufman (2008) found that about three quarters of studies used the *Torrance Tests for Creative Thinking*, making it difficult to evaluate any potential method effects. More recent studies that used other assessment techniques extending from behavioral to

neurological measures (e.g., Wallach-Kogan Creativity Tests: Cheung & Lau, 2010; diffusion weighted imaging: Ryman et al., 2014) showed differences in mean performance across gender groups and white matter connectivity in relation to creativity performance. Specifically, Cheung and Lau (2010) showed that a sample of early adolescent girls in Hong Kong performed better on verbal flexibility, figural fluency, figural flexibility, figural uniqueness, and figural unusualness on the Wallach-Kogan Creativity tests. Further, female college students involve more regions of the brain in producing novel ideas than male students, suggesting unique topographical organization of connectivity across genders (Ryman et al., 2014). These findings suggest that continued research that uses other assessments of creative ability may provide more insight into the differences in gender and creativity, if any do truly exist.

In addition to diverse measures of creative ability, it is important to consider potential domain differences. There are significant (although not large) mean gender differences in verbal ability and achievement (favoring girls and women; Halpern, 2000; Lewin et al., 2001; Petersen, 2018) and spatial ability (favoring boys and men; Lewin et al., 2001; Voyer et al., 2017). These differences could potentially be reflected in performance on creative thinking tasks. In the present study one of the performance criteria is use of emotions in creative storytelling (verbal domain) and drawing (figural domain) and previous research shows that girls have more sophisticated and granular emotion vocabularies (Bazhydai et al., 2019). Research on self-perceptions of abilities mirror (and magnify) observed differences in abilities, with males self-assessing their creativity as higher than the females' subjects did on science-analytic and sports areas and females rating themselves higher on social-communication and visual-artistic creative domains (Bender et al., 2013; Chan, 2005; Kaufman, 2006; Stoltzfus et al., 2011). Because creativity is at least in part domain specific (Baer, 1998, 2010; Ivcevic & Mayer, 2009), it is key

to examine gender differences in verbal and non-verbal domains, as well as in relation to different performance criteria.

Gender Variability and Creative Performance

A more recent line of research has focused on examining gender differences in variability on creative ability performance scores. A series of studies has aimed to investigate the Greater Male Variability Hypothesis (GMVH) in creative thinking abilities. This hypothesis stems from research showing greater male variability in cognitive ability and overrepresentation of males in both the lower and upper extremes of the distribution of scores (Feingold, 1992; Hedges & Friedman, 1993; Johnson et al., 2008). Greater male variability can be observed even when no mean gender differences are found across various areas of functioning, such as intellectual abilities (Feingold, 1992; Hedges & Nowell, 1995; Johnson et al., 2008). GMVH has been proposed to explain the tendency of men to dominate professional spheres and be overrepresented in high status positions (Noddings, 1992). By analogy, GMVH in creative abilities has been proposed to explain why males have greater creative achievement than females even though, when mean ability is examined, there seem to be no gender differences (He & Wong, 2011).

To examine differences in variability, two approaches are used. The first measure calculates variance ratios (VRs) – male variance in performance scored divided by female variance. A VR > 1.0 points to higher male variability and VRs < 1.0 indicates higher female variability. The second measure of variability analyzes the distribution of scores; performance scores are standardized and gender ratios in the different regions of the distribution – low scores, scores around the mean, and high scores – are examined (Hedges & Friedman, 1993). The

GMVH predicts that males are overrepresented in the low and high end of the distribution, while females are overrepresented around the mean.

The research conducted thus far on the GMVH in creative abilities has shown some mixed results. Initial evidence supported the hypothesis. He and Wong (2011) showed more boys in the lower and upper regions of the distribution, as well as VRs indicating greater male variability for 5th through 7th grade school children for the composite scores and 5 out of 10 TCT-DP performance criteria -- continuations, completions, connections with a line, boundary-breaking (dependent), and boundary-breaking (independent). Moreover, He (2018) found an overrepresentation of girls in the middle and upper regions of the distribution, and an overrepresentation of boys in the lower regions of the distribution when the participants were children. However, as adolescents, the female superiority was replaced by an overrepresentation of boys in both the lower and upper regions. A later study by Karwowski and colleagues (2016) found that, in a sample including children and adults (4-46 years of age), males across age groups displayed greater variability on originality and unconventionality and females showed greater variability on adaptiveness performance criteria on the TCT-DP.

However, other research shows varying support for the GMVH. He et al. (2013) did find that adolescent males had significantly greater variability than females, but only in the upper tails of the distribution of scores, not fully supporting the hypothesis. Ju et al. (2015) found no evidence for differences in variability on total scores in two separate school samples from urban and rural China on the TCT-DP.

Inconsistencies with the GMVH are more pronounced when creative performance is assessed across domains and tasks. For example, Lau and Cheung (2015) used the Wallach Kogan Creativity Tests (Wallach & Kogan, 1965) and found evidence for greater male

variability on figural divergent thinking tasks, but not on verbal divergent thinking tasks. Taylor and Barbot (2021) conducted two studies: one study evaluated performance of adults on a creative writing task, and another study evaluated performance of adolescents on a creative writing task and both figural and verbal divergent thinking tasks. Across both studies and all tasks used, there was no support for GMVH. There were more females in the lower region of the distribution for uncommonness on the verbal divergent thinking task, but no differences on other measures.

Two factors have been proposed to explain differences in findings about the GMVH in creativity – age and domain/task. Specifically, He (2018) followed children, adolescents, and young adults for four years and examined changes in variability of performance scores. In the child group, girls were overrepresented in the upper parts of the distribution and boys in the lower end of the distribution. There was a consistent overrepresentation of males in the lower region of the distribution across adolescence, and an overrepresentation of males in the upper region of the distribution starting at age 16. This pattern was also observed in young adults, pointing to possible gender socialization influences.

A large proportion of variance in performance on creative thinking tests can be described as task-specific (Barbot et al., 2016; Dumas & Dunbar, 2014). Although domain differences in male and female variability in creative ability can be expected theoretically, they are difficult to evaluate in previous research because most studies have only used the figural tests and largely relied on one specific task – TCT-DP. Because of this, it is not only important to study different domains of creative performance, but also include different tasks within specific creative domains.

Introduction to the Present Study

The present study examines gender differences in mean performance and variability of scores in a sample of primary and secondary school children. Gender differences in performance on verbal and figural creativity are analyzed using two tasks previously not employed – storytelling and incomplete figure drawing. Furthermore, performance is examined for different criteria, including creativity, novelty/originality, likeability, elaboration, humor, and emotion use.

Given that previous research on creative abilities suggest a relative gender equality and when gender differences exist, they tend to favor girls, combined with the strengths girls exhibit in verbal abilities (Bazhydai et al., 2019; Bender et al., 2013; Chan, 2005; Kaufman, 2006; Stoltzfus et al., 2011), if mean differences are found, we expect to find them for the storytelling task. Performance across the figural task is expected to be more mixed, with no one gender outperforming the other (Baer & Kaufman, 2008).

Previous research examining the GMVH suggests that there is a difference between childhood and early adolescence and later adolescence into adulthood (He et al., 2013; Ju et al., 2015; Lau & Cheung, 2015; Taylor & Barbot, 2021). The current study includes participants under the age of 16, thus if greater male variability emerges later in adolescence, we can hypothesize that in the present study girls will be overrepresented in the upper regions of the score distribution, while boys are overrepresented in the lower range of the score distribution.

Method

Sample

A total of 704 students, ranging from 1st-10th grade and ages 6-16 years old, attending 5 public schools in northern Spain participated in the current study. The schools drew students from a broad range of socio-economic and cultural backgrounds from the urban and suburban

environments. Because data for gender were missing for 26 students, the sample included in the analyses had a total of 678 students (337 boys, 49.7%; and 341 girls, 50.3%).

Measures

Figural creativity. An incomplete figures drawing task was adapted from a subtest of the Torrance Test of Creative Thinking (TTCT; Torrance, 1966). Participants were presented with an abstract shape comprised of three lines and given 3 minutes to complete a drawing using the shape in any way they desired, being told to tell a complete story with their picture. Participants were also asked to provide titles for their drawings.

Participants in grades 1-3 had instructions read to them by the experimenter in a one-on-one session. Participants in grades 4-6 had instructions read to them in a whole class setting, and participants in grades 7-10 independently read the instructions and completed the task in a whole class setting.

Two coders assigned a score of 1(lowest)-5(highest) for each of the 6 performance criteria: (1) creativity – how original and useful/appropriate the [drawing/story] was compared to others in the sample, (2) novelty/originality – how original or unique the [drawing and title/story] were compared to others in the sample, (3) likeability – how pleasing or enjoyable the [drawing/story] was compared to others in the sample, (4) elaboration – the amount of detail included in the [drawing/story], (5) humor – to what extent the artist included something funny or surprising in the [drawing/story], and (6) use of emotions – how much emotional content was included in the [drawing/story]. Titles/captions were translated from Spanish to English before coding. The inter-rater reliabilities were all within the acceptable to excellent range (creativity: $\alpha = .76$, novelty/originality: $\alpha = .80$, likeability: $\alpha = .70$, elaboration: $\alpha = .69$, humor: $\alpha = .66$, and emotion use: $\alpha = .68$). Scores were averaged to produce a single score for each variable.

Verbal creativity. Students completed a 5-10 minute storytelling task. Participants were presented with a series of pictures without words from the book, "A Boy, a Dog and a Frog" (Mayer, 1967), which depicts a short story about a boy and his dog trying to catch a frog in a river. Task administration was adapted to be age appropriate. For participants in 1st-3rd grade, the examiner presented the packet of pictures to each child and allowed them to look through the pictures to see what happens in the story. The child was then told that their job was to make up a story that goes along with the pictures and that the examiner would write down what they said. For participants in grades 4-6 (upper primary school), instructions were given to the entire class of students at once. They were instructed to look through all of the pictures and make up the story that goes along with the pictures. The students were then asked to write their responses on a provided form. For grades 7-10 (secondary school), students read the instructions silently, to themselves, before writing out their stories independently, with the examiner serving merely as the proctor for the class.

The Storytelling Task was coded using the same six categories as the Incomplete Figures Task: creativity, novelty/originality, likeability, elaboration, humor, and emotion use.
The coders were two Spanish-speaking research assistants who independently rated all of the stories, and inter-rater reliabilities were calculated for each of the variables. The alpha coefficients were all in the excellent range (creativity: $\alpha = .90$, novelty/originality: $\alpha = .91$, likeability: $\alpha = .94$, elaboration: $\alpha = .96$, humor: $\alpha = .91$, and emotion use: $\alpha = .91$). Ratings were averaged to provide a single score for each variable.

Procedure

Students were recruited to participate in the research study through letters and informed consent forms sent home to parents/legal guardians for review. If the consent forms were

returned to the school signed by parents/guardians, the student was eligible to participate. In addition, before beginning any of the assessments, the students were asked for assent.

Students completed assessments during school hours. Testing was conducted using a coding system to assure student anonymity and confidentiality and lasted less than thirty minutes.

Results

Three sets of analyses are presented: independent samples t-tests to compare mean performance for boys and girls, comparisons of boys/girls' variances, and comparisons of percentages of boys and girls in the different regions of the score distributions.

To examine gender differences in variances, we used Levene's test of homogeneity of variance and calculated variance ratios (VRs; boy's variance divided by girl's variance) as a measure of effect size. Feingold (1992, 1994) suggested that a VR > 1.10 or <0.90 should be considered the smallest meaningfully interpretable difference. However, tests of statistical significance for VRs meeting this suggested cut-off are inconsistent (Ju et al., 2015; Karwowski et al., 2016), pointing to the need to supplement tests of equality of variances with examination of percentages of boys and girls in the different portions of the distribution of scores.

The second set of measures examined proportions of boys and girls in the lower, middle, and upper regions of distribution of scores. Scores on different performance criteria were standardized and percentages of boys were divided by the percentages of girls in three groups: z-scores < -1, scores between -1 and 1, and scores > 1. Chi-square tests were used to assess statistical significance of differences in scores distributions (i.e. if one boys or girls were overrepresented in each of the three regions of the distribution). In addition to the chi-square

statistics, we report phi coefficients as a measure of effect size (interpreted as 0.10 being small, 0.30 medium, and 0.50 large effect).

Descriptive statistics by gender are presented in Table 1. In line with prior research, mean gender differences were not consistently found across all performance criteria and across tasks. However, when mean differences were significant, girls had higher scores, although all differences were small in size. This was the case for three of the six performance criteria on the figural task (likeability: t(627) = -3.15, p = .002, d = 0.25; elaboration: t(627) = -3.38, p < .001, d = 0.27; and emotion use: t(628) = -2.86, p = .004, d = 0.23) and for five of the six criteria on the verbal task (creativity: t(628) = -3.24, p = .001, d = 0.26; novelty: t(628) = -3.53, p < .001, d = 0.28; likeability: t(628) = -3.75, p < .001, d = 0.30; elaboration: t(628) = -4.15, p < .001, d = 0.33; and humor: t(624) = -2.35, p = .019, d = 0.19).

According to the Levene's test of equality of variances, there were significant gender differences in variability of scores for two performance criteria on the figural task (humor: F(1, 627) = 9.58, p = .002, VR = 0.68; emotion use: F(1, 627) = 12.31, p < .001, VR = 0.71) and two criteria on the storytelling task (novelty: F(1, 628) = 7.30, p = .007, VR = 0.93; humor: F(1, 628) = 5.27, p = .022, VR = 0.83). In each case, variability was greater for girls than boys.

Table 2 shows the percentages of boys and girls in the different regions of the distribution of performance criteria for the figural and verbal tasks, along with χ^2 test, and φ-coefficient. Overall, there was no evidence of greater male variability. For the figural task, there were more girls in the upper region of the distribution for likeability, $\chi^2(1) = 5.92$, $p = .015$, $\phi = .10$, and elaboration, $\chi^2(1) = 9.67$, $p = .002$, $\phi = .12$. There were more boys with average scores, and more girls in the upper region for humor and emotion use criteria, $\chi^2(1) = 5.13$, $p = .023$, $\phi = .09$ and $\chi^2(1) = 5.81$, $p = .016$, $\phi = -.10$, respectively.

Performance on the verbal storytelling task did not show support for the GMVH either, with differences in distribution differing across the performance criteria; creativity, more girls in the middle region, $\chi^2(1) = 6.30$, $p = .012$, $\phi = .10$; novelty, more girls in the lower region, $\chi^2(1) = 95.70$, $p < .001$, $\phi = .39$ and more boys in the middle region, $\chi^2(1) = 74.49$, $p < .001$, $\phi = .34$; likeability, more boys in the lower region, $\chi^2(1) = 5.80$, $p = .016$, $\phi = -.10$ and more girls in the upper region, $\chi^2(1) = 7.25$, $p = .007$, $\phi = .11$; elaboration, more boys in the lower region, $\chi^2(1) = 18.69$, $p < .001$, $\phi = -.17$ and more girls in the middle of the distribution, $\chi^2(1) = 6.11$, $p = .013$, $\phi = .10$; and emotion use, more boys in the lower end of the distribution, $\chi^2(1) = 3.92$, $p = .048$, $\phi = -.08$.

Discussion

The present study examined gender differences in performance on two creative thinking tasks – storytelling (task in the verbal domain) and completing line drawings (task in the figural domain) – in a sample of children and early to mid-adolescents. Gender differences in performance were examined both in relation to mean scores and variability in performance, specifically testing the greater male variability hypothesis. As in previous research, although mean performance differences were not found for all examined criteria, when such differences emerged, they favored girls (Baer & Kaufman, 2008). Greater male variability was not supported. Rather, there were more girls in the upper end of the distribution for several performance criteria across the two task domains and more boys in the lower end of the distribution for several criteria on the storytelling task.

Girls performed higher than boys for 3 out of 6 performance criteria on the figural task (likeability, elaboration, emotion use) and 5 of 6 criteria on the verbal task (all but emotion use). Where significant, differences in mean performance had small effect sizes ($0.19 < d < 0.33$).

Because girls tend to have higher verbal intelligence than boys (Halpern, 2000; Lewin et al., 2001), their advantage on storytelling could be due to the overall higher verbal ability. The limitations of the present study did not enable direct testing of this possibility. Although girls also tend to have more granular emotion vocabularies (Bazhydai et al., 2018), a mean difference in the use of emotions was not found for the storytelling. The reason for this finding could be that the measure of using emotions on this task was not sensitive to the sophistication or granularity in the ability to understand emotions and label them. Rather, a high score on the using emotion criterion indicated only the amount of emotion content.

Across both the verbal and figural tasks, girls had higher elaboration than boys. Elaboration captures the amount of detail included in performance and thus is akin to a measure of effort or persistence discussed by the dual process model of creative thinking (Nijstad et al., 2010). Gender differences in the tendency to solve creative thinking tasks through the flexibility or the persistence pathway has not been examined thus far. Because girls are higher on conscientiousness, personality dimension including traits such as persistence and hard work (Soto, 2016), the persistence pathway might be more likely in girls, contributing to their higher elaboration scores.

The present study examined two measures of variability – variance ratios and proportion of boys and girls in the different regions of score distributions. No evidence was found for the greater male variability hypothesis. Greater female variability was indicated by Levene's test of equality of variances for two criteria on the figural task (humor, emotion use) and two criteria for the storytelling task (novelty, humor). Examination of different portions of the distribution show that girls are overrepresented on these criteria in the upper end of scores. Although test of equality of variances did not show significant differences for other performance criteria, girls

were also overrepresented in the upper region of the distribution for likeability and elaboration on the figural task and likeability for the storytelling task. Boys were overrepresented in the lower end of the distribution on storytelling elaboration and emotion use and girls were overrepresented in the lower end of the distribution for storytelling novelty. Overall, the results do not show a consistent pattern of gender differences in score variability and when differences are observed, they tend to be small in size.

Although this study has strengths with regards to assessment of performance on verbal and figural tasks, it is important to note its limitations. This study included participants up to (approximately) 16 years of age. Because previous research suggested that age might play a role in the manifestation of greater male variability so that this aspect of gender differences only emerges in later adolescence (after the age of 16; He, 2018), future research should include participants in a broad range of ages, both children and early/mid adolescents as well as later adolescents and young adults. Such samples would enable a direct test of the age-related differences in variability of performance on creative thinking tests. While the present study did not provide evidence for greater male variability hypothesis, it is not possible to conclude whether this is due to the nature of the sample and age of participants or another factor (e.g., tasks specificity).

The possibility that findings about greater male variability are task-specific require more research. Barbot and colleagues (2016) showed that a substantial proportion of variance on tests of creative thinking is task-specific. In the context of gender differences in performance, it would be important to examine features of tasks on which variability in performance emerges. Because the present study included one verbal and one figural task, we are not able to address task vs. domain differences. To address these questions, researchers should aim to administer multi-

domain test batteries which provide multiple tasks per each domain, such as the EPoC (Barbot et al., 2016; Lubart et al., 2011; Lubart et al., 2013). Along with commonly studied verbal and figural domains, this battery also assesses creative thinking in social, scientific, math, music, and body movement domains. In addition to providing multiple tasks for each domain, EPoC also includes both measures of divergent-exploratory and convergent-integrative creative thinking.

It has been proposed that gender differences in creativity are in part due to gender differences in other traits or cognitive abilities. This proposition has not been previously explicitly addressed. Future studies would need to include a broad range of cognitive, self-concept, and personality assessments to start examining this theoretical proposition. For instance, it could be hypothesized that gender differences in mean performance and variability on verbal-literary creative thinking performance will only exist when relevant self-concept (e.g., self-evaluations of creativity for writing) and task-specific cognitive abilities (e.g., verbal intelligence) show gender differences.

Greater male variability has been proposed as offering (at least partial) explanation of gender differences in creative achievement in adulthood (He & Wong, 2011). Inconsistent findings about greater male variability at best puts into question this proposition. An ideal test of this proposition would include longitudinal studies where creative thinking abilities are tested in childhood or early adolescence (before emergence of hypothesized greater male variability), in late adolescence (as the hypothesized greater male variability emerges), and in (mid)adulthood (when creative achievement emerges). Practical challenges for such research are substantial and re-analysis of existing longitudinal studies could provide important insights. For example, in the late 1950s, Torrance started longitudinal studies of elementary and high school students, with several decades of follow up assessments (Cramond et al., 2005; Plucker, 1999; Runco et al.,

2010; Torrance, 1988). In addition to Torrance Tests of Creative Thinking, measures were collected for intelligence and adult public creative achievement (e.g., inventions, publications) and personal creative style of life achievements (e.g., designing a garden, organizing an action-oriented group).

Existing research suggests that any contribution of gender differences in performance on tests of creative thinking (whether mean or differences in variability) – if they reliably exist for some kinds of tasks – would be only a small part in explaining gender differences in creative achievement. Research shows that gender differences in achievement are related to social influences and biases, such as less support for creativity that women receive. Evidence is accumulating about the differential valuing of creativity of men and women in the workplace. Creative behavior is valued more for men than women, both based on correlational data in the workplace and experimental data examining evaluations of fictitious behavior (Luksyte et al., 2018?). Employees' self-reported creative behavior correlates with more positive performance evaluation ratings by supervisors and coworkers for men, but not for women. Similarly, manipulating gender of descriptions of creative work behavior showed that performance evaluations benefited from high creative behavior much more for male than for female employees. Similarly, research across industries in the US shows that this differential valuing of creative behavior is reflected in the subjective experience and behavior. Men report both more creative workplace behaviors (e.g., sharing creative ideas, contributing original ways to achieve goals) and greater support for creativity at work than women. Support for creativity at work mediates the relationship between gender and creative behavior. The proportion of women in an industry influences this relationship – differences between men and women become smaller with a greater proportion of women in an industry (Taylor et al., 2020). Social influence on creative

behavior and achievement is evident also from studies showing that although men have higher publicly recognizable creative achievement, personal creative behavior and achievement that is not dependent on social supports and criticisms does not show such gender-related differences (Runco et al., 2010).

Understanding gender differences in performance on creative thinking tests is an important question, both in terms of basic understanding of the nature of creative thinking abilities and potential implications for understanding real world creative behavior. Initial studies examining the greater male variability hypothesis seem to offer a relatively simple picture supporting the hypothesis, especially for the performance criteria pertaining to originality (He & Wong, 2011; Karwowski et al., 2016). As new tasks across different domains were administered to test the boundary conditions of gender-related variability, the picture became more complex and less clear (Taylor & Barbot, 2021). The present study adds to this complexity by examining the hypothesis in children and early adolescents and finds that in this sample, greater male variability is not supported. In line with previous research, when there are gender differences in mean performance, they favor girls.

The most important contribution of this paper is to pose more questions – questions of the role of tasks and domains and developmental differences in relation to performance on creativity tests. To address the prevalence of findings about greater male variability, future research will have to systematically assess multiple domains of creative performance, include multiple tasks per domain to identify task and domain effects and establish reliability of results. If studies with samples including a broad range of ages find that gender-related differences emerge in late adolescents, the questions of why will become key. One group of potential influences can include other cognitive abilities (e.g., verbal or spatial intelligence) and another group of

potential influences can include self-concept and social differences. Self-concept of traits and abilities reflects (to a significant extent) internalized societal stereotypes. Development of identity, including in relation to gender, in adolescence can be reflected in willingness to show nonconformity.

References

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. (1983). *The social psychology of creativity*. Springer-Verlag.

Baas, M., De Dreu, C. K., & Nijstad, B. A. (2008). A meta-analysis of 25 years of mood-creativity research: Hedonic tone, activation, or regulatory focus? *Psychological Bulletin*, *134*(6), 779-806. https://doi.org/10.1037/a0012815

Baer, J. (1993). *Creativity and divergent thinking: A task-specific approach*. Lawrence Erlbaum Associates.

Baer, J. (1998). The case for domain specificity of creativity. *Creativity Research Journal, 11*(2), 173–177. https://doi.org/10.1207/s15326934crj1102_7

Baer, J. (2010). Is creativity domain specific? In J.C. Kaufman & R.J. Sternberg (Eds.), *Cambridge Handbook of Creativity* (pp. 321–341). Cambridge University Press.

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

Barbot, B., Besançon, M., & Lubart, T. (2016). The generality-specificity of creativity: Exploring the structure of creative potential with EPoC. *Learning and Individual Differences*, *52*, 178-187. https://doi.org/10.1016/j.lindif.2016.06.005

Bazhydai, M., Ivcevic, Z., Brackett, M. A., & Widen, S. C. (2019). Breadth of emotion vocabulary in early adolescence. *Imagination, Cognition and Personality*, *38*(4), 378-404. https://doi.org/10.1177/0276236618765403

Bender, S. W., Nibbelink, B., Towner-Thyrum, E., & Vredenburg, D. (2013). Defining characteristics of creative women. *Creativity Research Journal*, *25*(1), 38-47. https://doi.org/10.1080/10400419.2013.752190

Chan, D. W. (2005). Self-perceived creativity, family hardiness, and emotional intelligence of Chinese gifted students in Hong Kong. *Journal of Secondary Gifted Education*, *16*(2-3), 47-56. https://doi.org/10.4219/jsge-2005-471

Cheung, P. C., & Lau, S. (2010). Gender differences in the creativity of Hong Kong school children: Comparison by using the new electronic Wallach–Kogan creativity tests. *Creativity Research Journal*, *22*(2), 194-199. https://doi.org/10.1080/10400419.2010.481522

Costa Jr, P. T., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: robust and surprising findings. *Journal of Personality and Social Psychology*, *81*(2), 322-331. https://doi.org/10.1037/0022-3514.81.2.322

Cramond, B., Matthews-Morgan, J., Bandalos, D., & Zuo, L. (2005). A report on the 40-year follow-up of the Torrance Tests of Creative Thinking: Alive and well in the new millennium. *Gifted Child Quarterly*, *49*(4), 283-291. https://doi.org/10.1177/001698620504900402

Dumas, D., & Dunbar, K. N. (2014). Understanding fluency and originality: A latent variable perspective. *Thinking Skills and Creativity*, *14*, 56-67. https://doi.org/10.1016/j.tsc.2014.09.003

Feingold, A. (1992). Sex differences in variability in intellectual abilities: A new look at an old controversy. *Review of Educational Research, 62,* 61–84. https://doi.org/10.3102/00346543062001061

Feingold, A. (1994). Gender differences in variability in intellectual abilities: A cross-cultural perspective. *Sex Roles, 30,* 81-92. https://doi.org/10.1007/bf01420741

Halpern, D. F. (2000). *Sex differences in cognitive abilities*. Psychology Press.

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

He, W. J., & Wong, W. C. (2011). Gender differences in creative thinking revisited: Findings from analysis of variability. *Personality and Individual Differences*, *51*(7), 807-811. https://doi.org/10.1016/j.paid.2011.06.027

He, W. J., Wong, W. C., Li, Y., & Xu, H. (2013). A study of the greater male variability hypothesis in creative thinking in Mainland China: Male superiority exists. *Personality and Individual Differences*, *55*(8), 882-886. https://doi.org/10.1016/j.paid.2013.07.017

Hedges, L. V., & Friedman, L. (1993). Gender differences in variability in intellectual abilities: A reanalysis of Feingold's results. *Review of Educational Research*, *63*(1), 94-105. https://doi.org/10.3102/00346543063001094

Hedges, L.V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science, 269,* 41–45. https://doi.org/10.1126/science.7604277

Hong, E., & Milgram, R. M. (2010). Creative thinking ability: Domain generality and specificity. *Creativity Research Journal*, *22*(3), 272-287. https://doi.org/10.1080/10400419.2010.503535

Ivcevic, Z., & Mayer, J. D. (2009). Mapping dimensions of creativity in the life-space. *Creativity Research Journal*, *21*(2-3), 152-165. https://doi.org/10.1080/10400410902855259

Johnson, W., Carothers, A., & Deary, I. J. (2008). Sex differences in variability in general intelligence: A new look at the old question. *Perspectives on Psychological Science*, *3*(6), 518-531. https://doi.org/10.1111/j.1745-6924.2008.00096.x

Ju, C., Duan, Y., & You, X. (2015). Retesting the greater male variability hypothesis in mainland China: A cross-regional study. *Personality and Individual Differences*, *72*, 85-89. https://doi.org/10.1016/j.paid.2014.07.021

Karwowski, M., Jankowska, D. M., Gralewski, J., Gajda, A., Wiśniewska, E., & Lebuda, I. (2016). Greater male variability in creativity: a latent variables approach. *Thinking Skills and Creativity*, *22*, 159-166. https://doi.org/10.1016/j.tsc.2016.10.005

Karwowski, M., Lebuda, I., Wisniewska, E., & Gralewski, J. (2013). Big five personality traits as the predictors of creative self-efficacy and creative personal identity: Does gender matter? *The Journal of Creative Behavior*, *47*(3), 215-232. https://doi.org/10.1002/jocb.32

Kaufman, J. C. (2006). Self-reported differences in creativity by ethnicity and gender. *Applied Cognitive Psychology*, *20*(8), 1065-1082. https://doi.org/10.1002/acp.1255

Kaufman, J. C., Baer, J., Agars, M. D., & Loomis, D. (2010). Creativity stereotypes and the consensual assessment technique. *Creativity Research Journal*, *22*(2), 200-205. https://doi.org/10.1080/10400419.2010.481529

Kaufman, J. C., Baer, J., & Gentile, C. A. (2004). Differences in gender and ethnicity as measured by ratings of three writing tasks. *The Journal of Creative Behavior*, *38*(1), 56-69. https://doi.org/10.1002/j.2162-6057.2004.tb01231.x

Kim, K. H. (2005). Can only intelligent people be creative? A meta-analysis. *Journal of Secondary Gifted Education*, *16*(2-3), 57-66. https://doi.org/10.4219/jsge-2005-473

Kling, K. C., Hyde, J. S., Showers, C. J., & Buswell, B. N. (1999). Gender differences in self-esteem: a meta-analysis. *Psychological Bulletin*, *125*(4), 470-500. https://doi.org/10.1037/0033-2909.125.4.470

Kogan, N. (1974). Creativity and sex differences. *Journal of Creative Behavior, 8,* 1-14. https://doi.org/10.1002/j.2162-6057.1974.tb01103.x

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

Lau, S., & Cheung, P. C. (2015). A gender-fair look at variability in creativity: Growth in variability over a period versus gender comparison at a time point. *Creativity Research Journal*, *27*(1), 87-95. https://doi.org/10.1080/10400419.2015.992685

Lebuda, I., Zabelina, D. L., & Karwowski, M. (2016). Mind full of ideas: A meta-analysis of the mindfulness–creativity link. *Personality and Individual Differences*, *93*, 22-26. https://doi.org/10.1016/j.paid.2015.09.040

Lewin, C., Wolgers, G., & Herlitz, A. (2001). Sex differences favoring women in verbal but not in visuospatial episodic memory. *Neuropsychology*, *15*(2), 165-173. https://doi.org/10.1037/0894-4105.15.2.165

Lubart, T., Besancon, M., & Barbot, B. (2011). *Evaluation du Potentiel Creatif (EPoC).* Hogrefe.

Lubart, T., Zenasni, F., & Barbot, B. (2013). Creative potential and its measurement. *International Journal for Talent Development and Creativity*, *1*(2), 41-50. https://doi.org/10.4135/9781412971959.n292

Luksyte, A., Unsworth, K. L., & Avery, D. R. (2018). Innovative work behavior and sex-based stereotypes: Examining sex differences in perceptions and evaluations of innovative work behavior. *Journal of Organizational Behavior*, *39*(3), 292-305. https://doi.org/10.1002/job.2219

Mayer, M. (1967). *A boy, a dog, and a frog.* Dial Books.

Misra, I. (2003). Openness to experience: Gender differences and its correlates. *Journal of Personality and Clinical Studies*, *19*(1), 141-152. https://doi.org/10.1016/j.paid.2014.02.012

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

Noddings, N. (1992). Variability: A pernicious hypothesis. *Review of Educational Research*, *62*(1), 85-88. https://doi.org/10.3102/00346543062001085

Pagnani, A. R. (2011). Gender differences. In Editors-in-Chief: M. A. Runco & S. R. Pritzker (Eds.), *Encyclopedia of creativity (second edition)* (pp. 551–557). Academic Press.

Petersen, J. (2018). Gender difference in verbal performance: A meta-analysis of United States state performance assessments. *Educational Psychology Review, 30*(4), 1269–1281. https://doi.org/10.1007/s10648-018-9450-x

Plucker, J. A. (1999). Is the proof in the pudding? Reanalysis of Torrance's longitudinal data. *Creativity Research Journal, 12*, 103–114. https://doi.org/10.1207/s15326934crj1202_3

Runco, M. A., Cramond, B., & Pagnani, A. R. (2010). Gender and creativity. In J. C. Chrisler & D. R. McCreary (Eds.), *Handbook of gender research in psychology* (pp. 343–357). Springer.

Runco, M. A., Millar, G., Acar, S., & Cramond, B. (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

Ryman, S. G., van den Heuvel, M. P., Yeo, R. A., Caprihan, A., Carrasco, J., Vakhtin, A. A., Flores, R. A., Wertz, C., & Jung, R. E. (2014). Sex differences in the relationship between white matter connectivity and creativity. *NeuroImage, 101,* 380-389. https://doi.org/10.1016/j.neuroimage.2014.07.027

Sayed, E. M., & Mohamed, A. H. H. (2013). Gender differences in divergent thinking: use of the test of creative thinking-drawing production on an Egyptian sample. *Creativity Research Journal*, *25*(2), 222-227. https://doi.org/10.1080/10400419.2013.783760

Soto, C. J. (2016). The Little Six personality dimensions from early childhood to early adulthood: Mean-level age and gender differences in parents' reports. *Journal of Personality, 84*(4), 409–422. https://doi.org/10.1111/jopy.12168

Stoltzfus, G., Nibbelink, B. L., Vredenburg, D., & Thyrum, E. (2011). Gender, gender role, and creativity. *Social Behavior and Personality: An International Journal*, *39*(3), 425-432. https://doi.org/10.2224/sbp.2011.39.3.425

Taylor, C. L., & Barbot, B. (2021). Gender differences in creativity: Examining the greater male variability hypothesis in different domains and tasks. *Personality and Individual Differences, 174,* https://doi.org/10.1016/j.paid.2021.110661

Taylor, C., Ivcevic, Z., Moeller, J., & Brackett, M. (2020). Gender and support for creativity at work. *Creativity and Innovation Management*, *29*(3), 453-464. https://doi.org/10.1111/caim.12397

Torrance, E. P. (1966). The Torrance Tests of Creative Thinking–Norms—Technical Manual Research Edition—Verbal Tests, Forms A and B—Figural Tests, Forms A and B. Personnel Press.

Torrance, E. P. (1988). The nature of creativity as manifest in its testing. In R. J. Sternberg (Ed.), *The nature of creativity* (pp. 43-75). Cambridge University Press.

Ulger, K. (2014). The investigation of the students' creative thinking development. *Education and Science, 39,* 275-284. https://doi.org/10.15390/eb.2014.2160

Urban, K. K. (1991). On the development of creativity in children, *Creativity Research Journal*, 4(2), 177- 191. https://doi.org/10.1080/10400419109534384

Urban, K. K., & Jellen, H. G. (1996). Manual of Test for Creative Thinking – Drawing Production (TCT-DP). Harcourt Test.

Voyer, D., Voyer, S. D., & Saint-Aubin, J. (2017). Sex differences in visual-spatial working memory: A meta-analysis. *Psychonomic Bulletin & Review, 24*(2), 307–334. https://doi.org/10.3758/s13423-016-1085-7

Wallach, M. A., & Kogan, N. (1965*). Modes of thinking in young children: A study of the creativity–intelligence distinction.* New York: Holt, Rinehart, & Winston.

Table 1

Descriptive statistics by gender, mean differences, and variability ratios

	Male			Female			Mean diference	Variability	
	M	Var	Min-Max	M	Var	Min-Max	t	VR	Levene's test
Figural task									
Creativity	2.88	0.90	1-5	2.93	0.99	1-5	-0.62	0.91	0.85
Novelty	3.37	1.26	1-5	3.28	1.31	1-5	0.16	0.86	0.61
Likeability	2.69	0.95	1-5	2.95	1.11	1-5	-3.15***	0.96	1.96
Elaboration	2.82	0.81	1-5	3.07	0.90	1-5	-3.38***	0.90	2.47
Humor	1.28	0.24	1-3.5	1.35	0.35	1-4	-1.56	0.68	9.58**
Emotion	1.57	0.66	1-4.5	1.77	0.93	1-5	-2.86**	0.71	12.31***
Storytelling									
Creativity	2.54	1.34	1-5	2.84	1.41	1-5	-3.24***	0.95	0.06
Novelty	1.99	1.10	1-5	2.29	1.18	1-5	-3.53***	0.93	7.30**
Likeability	2.37	1.29	1-5	2.71	1.38	1-5	-3.75***	0.93	0.24
Elaboration	2.16	1.19	1-5	2.53	1.27	1-5	-4.15***	0.94	0.92
Humor	1.63	1.16	1-5	1.85	1.39	1-5	-2.35*	0.83	5.27*
Emotion	2.29	1.30	1-5	2.42	1.22	1-5	-1.53	1.06	0.01

Note. $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .010$.

Table 2

Male and female participants in different regions of the distribution on incomplete figures and storytelling tasks.

	Boys		Girls		Boy/girl ratio	χ^2	φ
	n	% within gender	n	% within gender			
Figural task							
Creativity							
z < -1	37	11.9	43	13.5	0.88	0.37	.02
-1 < z < 1	219	70.4	206	64.8	1.09	2.28	-.06
z > 1	55	17.7	69	21.7	0.82	1.60	.05
Novelty							
z < -1	67	21.5	75	23.6	0.91	0.38	.02
-1 < z < 1	161	51.8	161	50.6	1.02	0.08	-.01
z > 1	83	26.7	82	25.8	1.03	0.07	-.01
Likeability							
z < -1	62	19.9	55	17.3	1.15	0.72	-.03
-1 < z < 1	198	63.7	186	58.5	1.09	1.77	-.05
z > 1	51	16.4	77	24.2	0.68	5.92*	.10
Elaboration							
z < -1	90	28.9	72	22.6	1.28	3.26	-.07
-1 < z < 1	168	54.0	159	50.0	1.08	1.02	-.04
z > 1	53	17.0	87	27.4	0.62	9.67**	.12
Humor							
z < -1	0	0	0	0	NA		
-1 < z < 1	269	86.5	252	79.2	1.09	5.81*	-.10
z > 1	42	13.5	66	20.8	0.65	5.81*	.10
Emotion							
z < -1	0	0	0	0	NA	NA	NA
-1 < z < 1	276	88.7	262	82.4	1.08	5.13*	-.09
z > 1	35	11.3	56	17.6	0.64	5.13*	.09

Storytelling							
Creativity							
z < -1	58	18.5	46	14.5	1.28	1.85	-.05
-1 < z < 1	200	63.9	232	73.2	0.87	6.30*	.10
z > 1	55	17.6	39	12.3	1.43	3.44	-.07
Novelty							
z < -1	0	0	84	26.5	NA	95.70***	.39
-1 < z < 1	285	91.1	196	61.8	1.47	74.49***	.34
z > 1	28	8.9	37	11.7	0.76	1.26	.04
Likeability							
z < -1	80	25.6	56	17.7	1.44	5.80*	-.10
-1 < z < 1	183	58.5	183	57.7	1.01	0.04	-.01
z > 1	50	16.0	78	24.6	0.65	7.25***	.11
Elaboration							
z < -1	108	34.5	61	19.2	1.80	18.69***	-.17
-1 < z < 1	163	52.1	196	61.8	0.84	6.11*	.10
z > 1	42	13.4	60	18.9	0.71	3.52	.08
Humor							
z < -1	0	0	0	0	NA	NA	NA
-1 < z < 1	268	85.6	280	88.3	0.97	1.02	.04
z > 1	45	14.4	37	11.7	1.23	1.02	-.04
Emotion							
z < -1	96	30.7	75	23.7	1.30	3.92*	-.08
-1 < z < 1	169	54.0	195	61.5	0.88	3.65	.08
z > 1	48	15.3	47	14.8	1.03	0.03	-.01

Note. $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .010$.