**Gender Differences in Creativity:**

**Examining the Greater Male Variability Hypothesis in Different Domains and Tasks**

Christa L. Taylor1 and Baptiste Barbot1,2

1 Université catholique de Louvain

2 Yale University

Christa L. Taylor, https://orcid.org/0000-0002-6472-3107

Baptiste Barbot, https://orcid.org/0000-0002-5096-2596

The authors have no conflicts of interest to disclose. The preparation of this article was supported by grant RFP-15-05 to Baptiste Barbot from the Imagination Institute (www.imagination-institute.org), funded by the John Templeton Foundation, and the 2019-2022 Francqui Foundation start-up grant to Baptiste Barbot. This research was partly supported by the Spencer Foundation (Grant #201200103, PI: Baptiste Barbot). The opinions expressed in this publication are those of the authors and do not necessarily reflect the view of the funding agencies. We thank Alexandra Blanchard, Brianna Heuser, Rosemary Leo, Catalina Mourgues, Danielle Saliani, Mansi Shah, Baruch Wolhendler, and Silvia Hein for help with data collection and Lihong Xie for rating stories for the first study.

Correspondence concerning this article should be addressed to Baptiste Barbot, Université catholique de Louvain, IPSY Place Cardinal Mercier 10/L3.05.01, 1348 Louvain-la-Neuve, Belgium. Email: baptiste.barbot@uclouvain.be

**Abstract**

The Greater Male Variability Hypothesis (GMVH) suggests that males demonstrate greater variability than females and are overrepresented in the lowest and highest ranges of cognitive ability. Several studies have found evidence for the GMVH in creative performance, yet nearly all have used the same task (i.e., Test of Creative Thinking Drawing-Production) limiting inference to a single domain and modality of creative ability. In two studies, we examine the GMVH in relation to performance by adults (Study 1; *N* = 120) on a creative writing task and by adolescents (Study 2; *N* = 529) on a creative drawing task, as well as figural and verbal divergent thinking tasks. The variability of scores did not differ substantively between males and females on any of the tasks in either study and the pattern of the proportions of males and females in different regions of the distribution of scores was inconsistent across tasks. Although males received significantly greater mean scores than females on the verbal divergent thinking task in Study 2, no significant mean differences were found for any other task. Overall, our results do not support the GMVH and suggest that, if any, gender differences in creative variability are likely inconsistent across domains and tasks.

***Keywords: ﻿***Creativity; Divergent Thinking; Gender; Greater Male Variability Hypothesis

**Gender Differences in Creativity:**

**Examining the Greater Male Variability Hypothesis in Different Domains and Tasks**

Despite much research on gender and creativity, there is little evidence for gender differences on measures of creative potential and ability (Abraham, 2016; Baer & Kaufman, 2008; Runco et al., 2010). Most studies examining gender differences in creative ability have focused on divergent thinking (i.e., the ability to generate many original responses to a single stimulus; Baer & Kaufman, 2008; Guilford, 1956), which has long been considered a key indicator of creative *potential* (Runco & Acar, 2012). Although the majority of these studies show no difference between men and women, when a difference is found, women tend to score higher than men (see Baer & Kaufman, 2008). However, there are pronounced differences in the creative *achievement* of men and women, with men being overrepresented at higher levels of achievement (Simonton, 1992, 1994). It is currently not clear why this gender gap in creativity exists and, specifically, why women exhibit similar (or higher) levels of creative potential as men but lag behind in creative achievement. Several authors have suggested that the difficulty women face in creative pursuits may be due in part to more limited resources and opportunities, greater family and relationship demands, and/or overt opposition from a male-dominated society (Runco et al., 2010; Simonton, 1994).

One explanation for gender differences in creative achievement suggested in recent years is based on the Greater Male Variability Hypothesis (GMVH; He & Wong, 2011). The GMVH, which suggests that males demonstrate greater variability in cognitive ability than females (Ellis, 1894/1934), stems from research on intelligence and related cognitive abilities (Feingold, 1992; Johnson et al., 2008). Some evidence suggests that, on tests of cognitive ability (e.g., nonverbal reasoning and spatial ability), males demonstrate greater variability in scores than females and are overrepresented in both the lower and upper extremes of the distribution of scores (Feingold, 1992; Hedges & Friedman, 1993; Hedges & Nowell, 1995; Machin & Pekkarinen, 2008). Although the GMVH was historically explained using biological factors that positioned women as innately undifferentiated beings, modern researchers have suggested both biological and socio-cultural explanations (e.g., gender-based differences in expectations for social conformity; see Noddings, 1992). Gender differences in variability are classically determined using male/female variance ratios (VR; i.e., male variance divided by female variance), wherein a VR greater than one indicates greater male variability. The proportion of men and women in different regions of the distribution of scores is typically examined using the ratio of male/female participants in the mid-range, upper, and lower tails of the distribution. According to the GMVH, females are overrepresented (i.e., M/F ratio < 1) in the mid-range of a distribution of scores, whereas males are overrepresented (i.e., M/F ratio > 1) in the extremes. The overrepresentation of males in the upper tail of the distribution of cognitive abilities has been suggested to explain the greater prevalence of eminent men (compared to women) since the early 1900s (Noddings, 1992). Several studies have demonstrated support for the GMVH in creativity and suggested that the overrepresentation of males in the upper tail of the distribution may explain why men demonstrate greater real-world creative achievement than women in the absence of mean differences in creative potential (e.g., He & Wong, 2011).

**Greater Male Variability in Creativity**

In a seminal study of the GMVH in creativity, He and Wong (2011) administered the Test for Creative Thinking-Drawing Production (TCT-DP; Urban & Jellen, 1996) to Chinese school children. The TCT-DP asks participants to create a drawing using a sheet of paper containing six figural fragments. The drawing is then scored on 14 dimensions purported to reflect creativity from a gestalt perspective (e.g., using symbols or signs, connecting the figural fragments, breaking away from two dimensions; Urban, 2004). No mean differences were found between boys and girls on the total scale score (He & Wong, 2011). However, boys’ scores demonstrated significantly greater variability than girls — as established by a VR above 1.00 and significant ﻿*F* test of equality of variance —, and boys were overrepresented in the upper and lower tails of the distribution of scores. Greater male variability on the TCT-DP has been replicated with adult Meru tribe members in Kenya, Africa (Karwowski, Jankowska, Gajda, et al., 2016) and Polish children and adults (Karwowski, Jankowska, Gralewski, et al., 2016).

However, several studies have found results inconsistent with the GMVH in creativity (He et al., 2013; Ju et al., 2015). He et al. (2013) found that adolescent males had significantly higher mean scores on the total TCT-DP and demonstrated significantly greater variability than females, yet males were overrepresented only in the upper tails of the distribution of scores. Ju et al. (2015) found that results differed for school children from urban and rural China. Although boys from urban China received significantly greater mean creativity scores than girls, boys from rural China did not significantly differ from girls, and there was no evidence for differences in variability on total scale scores in either sample. When examining the distributions, urban boys were significantly overrepresented only in the upper tail of the distribution and rural boys were overrepresented only in the lower tail of the distribution. Thus, there are discrepancies in studies examining the GMVH in creativity.

He (2018) suggested that discrepancies in the literature examining gender differences in creativity may be due to differences in the age groups assessed, or to collapsing scores across age groups. In a four-year longitudinal study, He (2018) administered the TCT-DP annually to Chinese participants in four age groups (ranging from 8 to 18 years old at time 1). Female participants demonstrated greater mean scores than male participants in groups under the age of 16, whereas mean scores did not differ by gender for those over the age of 16. Additionally, male participants were significantly overrepresented in the lower tails of the distribution of scores across all age groups and overrepresented in the upper tails of the distribution for groups over the age of 16. Thus, He (2018) suggested that inconsistent findings for the GMVH may stem from the developmental patterns of gender differences in creativity. However, studies examining the GMVH in creativity have only used the TCT-DP to assess aspects of creativity (with the exception of Lau & Cheung, 2015). It is questionable whether broad generalizations about the nature of creativity or its development can be made based on performance on a single task corresponding with one modality (Barbot et al., 2019), calling for more studies investigating the GMVH with a more diverse set of creative tasks.[[1]](#footnote-2)

**Creative Domain and Task Specificity**

Although creative ability has often been conceptualized as domain-general (i.e., creative ability is the same across different domains; Plucker, 1998), most modern creativity conceptualizations acknowledge some level of domain-and task-specificity (i.e., creative ability varies from domain to domain or task to task within the same domain; Baer, 1998, and even from stimulus to stimulus within the same task; Barbot, 2018). This is due in part to the development of several theories, and supporting evidence, delineating how creativity requires both domain general and domain specific resources (Baer & Kaufman, 2005; Plucker et al., 2004). For instance, the Amusement Park Theoretical model of creativity (APT; Baer & Kaufman, 2005, 2017) suggests that although there are general requirements for all creative behavior (e.g., minimum levels of intelligence and motivation to engage in creative behavior), creative outcomes differ based on domain-relevant skills and characteristics. Indeed, studies that focus on general characteristics of a creative person (e.g., openness to experience) are more likely to support a domain-general view, whereas studies that focus on creative performance (e.g., based on ratings of creative products) are more likely to support a domain-specific view (Plucker, 1999, 2004; Qian & Plucker, 2018).

Performance in different creative domains (i.e., areas in which people engage in creativity) relates to various factors in distinct ways. Research using factor analysis to identify the organization of peoples’ self-rated creative perceptions and achievements across a range of tasks most often suggests three to five factors (Carson et al., 2005; Kaufman, 2012; Kaufman et al., 2009; Kaufman & Baer, 2004; McKay et al., 2016; Rawlings & Locarnini, 2007; Silvia et al., 2012). For instance, Kaufman and colleagues (Kaufman, 2012; McKay et al., 2016) found evidence for five distinct creative domains: ﻿everyday, scholarly, performance, science, and the arts. Creative products from different domains (e.g., stories, drawings, mathematical equations) created by the same individuals are infrequently correlated or are only weakly correlated with one another (Amabile, 1996; Baer, 1993, 2015; Barbot, 2020; Conti et al., 1996; Taylor & Kaufman, 2020). Additionally, various skills, thinking processes, and personality factors have been found to relate to creativity across domains and tasks in distinct ways (Barbot, 2020; Barbot et al., 2016; Jeon et al., 2011; Ruscio et al., 1998; Taylor & Kaufman, 2020).

For example, Ruscio et al. (1998) found that striving was a significant predictor of verbal creativity, but not visual creativity or problem-solving, and Barbot (2020) found that global self-esteem significantly predicted performance on a creative musical task, but not literary-verbal or graphic tasks. Furthermore, Barbot et al. (2016) found that domain-general variance accounted for only up to 11.7% – whereas, domain-specific skills accounted for up to 59.1% and task-specific skills accounted for up to 46.3% – of the variance in scores on a range of figural and verbal creativity tasks.

There is reason to believe that gender differences in creative variability may also differ across domains and tasks. Lau and Cheung (2015) analyzed normative data for Chinese school children from the Wallach-Kogan Creativity Tests (Wallach & Kogan, 1965) and found evidence for greater male variability on figural divergent thinking tasks across most age groups. However, no evidence for gender differences in creative variability was found for verbal divergent thinking tasks. These results are similar to findings in the literature regarding the GMVH in intelligence and other cognitive abilities, wherein greater male variability was found on spatial, but not verbal, tasks (e.g., Feingold, 1992). Additionally, the criteria on which the TCT-DP is scored differs substantially from other tasks commonly used to assess creativity (Urban, 2004). For instance, the Consensual Assessment Technique (CAT; Amabile, 1982) is a “gold standard” of creativity assessment (Baer & McKool, 2014) and consists of products being subjectively rated for creativity by a group of experts in a given domain (see also Cseh & Jeffries, 2019). Despite its popularity, there has been no evidence provided for the GMVH using CAT-rated products.

**The Present Studies**

We conducted secondary analyses of data from two studies to address the possibility that gender differences in creativity variability may differ across domains and tasks. Study 1 examines the GMVH in the creative writing of adults. If greater male variability in creativity can be generalized to domains/task modalities other than that assessed by the TCT-DP and is exhibited primarily after early adolescence (e.g., He, 2018), men should also exhibit greater variability in creativity scores than women on such a task. Study 2 assesses creativity in drawing, as well as verbal and figural divergent thinking tasks, in adolescents. In general, we expect that gender differences in creative variability will differ by domain and task. Consistent with Lau and Cheung (2015), and mirroring findings for GMVH on spatial versus verbal cognitive abilities (e.g., Feingold, 1992), we expect that males will demonstrate greater variability on the figural tasks, but not the verbal tasks. Together, these studies seek to illuminate the generalizability of the GMVH to different creative tasks and modalities, with adults and adolescents in the United States.

**Study 1**

**Method**

***Participants***

This study was approved by the Institutional Review Board at (approval #redacted) and different analyses using this data have been reported in Redacted for blind review and Redacted for blind review. Participants were adults recruited from a University in the Northeastern United States using flyers, in-class announcements, and events. All participants had scores lower than a clinical threshold on the Prodromal Questionnaire-brief version (Loewy et al., 2011). Data for those not reporting either male or female gender were excluded from analyses, resulting in a sample of 120 participants (63.3% female, 36.7% male). Participants’ ages ranged from 18 to 53 years old (*M* = 21.38, *SD* = 4.45). Participants’ education level ranged from completing high school to earning a graduate degree, with the majority indicating some college (60.8%) or receiving a Bachelor’s degree (20%). Race and Ethnicity were represented as follows: 47.5% Caucasian/White, 20.8% Multiracial, 15.8% Asian, 15% African American/Black, and .8% American Indian/Alaska Native. 25.8% of participants identified as Hispanic. Men and women did not differ significantly by age, *t* (118) = .33, *p* = .74, *d* = .07, 95% CI [-.31, .44], education level, χ2 (4) = 5.26, *p* = .26, *ϕc =* .21, or ethnic representation, *χ*2 (4) = 4.65, *p* = .33., *ϕc =* .20. All participants had the opportunity to be included in a raffle to win one of three $50 gift cards, and participants who completed additional portions of the study were also compensated with a $25 gift card.

***Materials***

**Storyboard Task.** Creativity was assessed using the storyboard task (Taylor et al., 2020), a creative writing assessment based on the interpretation and integration of visual information. Each storyboard item consists of three related black and white photographs presented as the beginning, middle, and end of a story along with three separate text boxes corresponding to each. Participants are asked to *“Write a story using each picture as an illustration of the beginning, middle, and end of your story.”* Participants completed two forms (A and B), each consisting of two storyboard items. Both forms were completed by 70% of participants (13% completed only form A and 17% completed only form B). Responses were rated using the CAT (Amabile, 1982), on a scale from 1 (not creative at all) to 7 (extremely creative) by three creativity researchers. Inter-rater reliability was acceptable for all four stories: item A1 (α = .82), A2 (α = .85), B1 (α = .79), and B2 (α = .75).[[2]](#footnote-3) Mean creativity scores were calculated by averaging ratings across raters for each storyboard item. The consistency of CAT scores across the four items was acceptable (α = .79). Therefore, scores across the four items were averaged (using only available data for cases with missing values) to create a total mean creativity score for each participant.

***Procedure***

This study was approved by the Institutional Review Board at the participating university. As part of a larger study, participants completed a battery of creativity-related assessments in the lab and at home across five weeks (see redacted for blind review). The storyboard task was completed on a tablet device with a keyboard during the first and last in-lab session, with forms A and B counterbalanced. Participants viewed an example storyboard task and instructions before beginning and the tasks were self-paced.

***Data Analyses***

Consistent with several studies examining the GMVH in intelligence and creativity (e.g., Arden & Plomin, 2006; Karwowski, Jankowska, Gajda, et al., 2016), we used Levene’s test of homogeneity of variance to test gender differences in creativity variability[[3]](#footnote-4) and calculated variance ratios (VR) as an effect size. VRs were calculated by dividing male variance by female variance, wherein VRs greater than 1.00 indicate greater male variability and less than 1.00 indicate greater female variability. Feingold (1992, 1994) suggested that, in descriptive terms, a VR > 1.10 (or < .90) —i.e., corresponding to 10% difference— may be considered the smallest meaningful effect size. In other words, the variances between groups may be considered equal for VRs between .90 and 1.10. Indeed, because VRs are calculated from sample variances, they will likely deviate from 1.00 (even if population variances are equal) due to normal sampling error (Ruscio & Roche, 2012). In the GMVH of creativity, a VR = 1.10 means that the variance for males is 10% greater than the variance for females, which would be statistically significant with about 1,000 participants in each group being compared (Feingold, 1992). However, no justification has been provided for the meaningfulness of a 10% difference in variability.

The size of variance ratio that denotes a meaningful difference in variability between groups is debatable and has direct implications for the interpretation of the significance of Levene’s test. A recent simulation study suggests that 120 participants are needed for 80% power to detect a VR = 1.5 using Levene’s test (assuming equal group sizes; Delacre et al., 2017). However, the most conservative commonly accepted rule of thumb suggests that variance homogeneity can be assumed with a VR ≤ 3.00 (when dividing the larger variance by the smaller; e.g., Dean & Voss, 1999; Keppel et al., 1992; Kirk, 2013). Ruscio and Roche (2019) found that, across randomly selected studies from ﻿diverse subfields within psychology, the mean VR from 234 studies comparing two groups was 2.51. Additionally, because Levene’s test is sensitive to sample size, the statistical significance associated with different VRs found in previous studies examining the GMVH in creativity has been inconsistent. For instance, in two previous studies, both reporting a VR = 1.17, Levene’s test was statistically significant in one case (*N* = 5,526; Karwowksi, Jankowska, Gralewski et al., 2016), yet not in the other case (*N* = 515; Ju et al., 2015). Thus, the results of Levene’s test alone should not be interpreted as evidence for or against the GMVH, but rather, the latter should rely on the broader pattern of results, including the VR and the ratio of males and females in different regions of the distribution.

The ratio of males and females in different regions of the distribution of creativity scores (M/F ratio) were computed by dividing the number of males (percent within gender to control for differences between the groups in sample size) by females in each region of the distribution of scores. M/F ratios were calculated for the regions (1) 1 SD below the mean, (2) between -1 SD and 1 SD, and (3) 1 SD above the mean. Although M/F ratios were calculated for comparison with previous studies examining the GMVH in creativity, it is important to note that they are descriptive and are not to be interpreted as VRs. Chi-square tests were used to determine if one gender was significantly overrepresented in different regions of the distributions. *A priori* power analysis estimated that, for 80% power to detect a small, medium, or large effect size with the chi-square test, one requires 785, 88, or 32 participants, respectively (using G\*Power 3.1; Faul, Erdfelder, Buchner, & Lang, 2009). Although this means that several of the tests may be underpowered in the current study given the data available for re-analysis, this situation is not uncommon in GMVH studies as sample sizes are, by-design, limited in the extreme regions of the distribution; Previous studies examining the GMVH in creativity with larger total sample sizes have conducted the test with similarly low numbers of participants in the tails of a distribution. For example, He and Wong’s (2011) study of 985 participants reports results for this test for *n* = 36 participants and Ju et al.’s (2015) study of 515 rural participants reports results for this test for *n* = 6 participants in the tails of the distributions. Therefore, although results for the chi-square tests are reported in the current study, the phi coefficient (*ϕ*) is also reported as a measure of effect size, which may be interpreted using Cohen's (1992) guidelines for *r*, wherein *.*10 is small,.30 is medium, and .50 is large.

**Results**

The distributions of creativity scores by gender did not deviate substantially from normal and no univariate outliers (*z* > ± 3.5 SD from the mean) were detected. The descriptive statistics for Studies 1 and 2 may be seen in Table 1. Mean creativity on the storyboard task did not differ significantly between men and women, *t*(118) = -1.45, *p* = .15, *d* = -.27, 95% CI [-.65, .09].[[4]](#footnote-5) There was also no significant gender difference in the variability of the creativity scores, according to Levene’s test: *F*(1, 118) = 2.67, *p* = .11, with a VR = 1.56.

The total distribution of creativity scores did not deviate from normal (Skew = .28, SE = .22; Kurtosis = .10, SE = .44). The M/F ratios, calculated using the percentage of men and women within gender group that fell within the three regions of the distribution of creativity scores, may be seen in Table 2. Although the within-gender group percentage of women was slightly higher than that for men in the lower tail (more than 1 SD below the mean) and mid-range (within 1 SD below and above the mean) of the distribution of creativity scores, the M/F ratios were not significant. However, men were significantly overrepresented in the upper tail of the distribution (i.e. scores higher than 1 SD above the mean), *χ*2 (1) = 4.38, p < .05, *ϕ* = .48.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 1**  *Descriptive statistics by gender for Study 1 and 2* | | | | | | | | | | | | | |
|  | Male | | | | | |  | Female | | | | | |
|  | *N* | M | Var | Min-Max | Skew | Kurtosis |  | *N* | M | Var | Min-Max | Skew | Kurtosis |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Study 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Storyboard | 44 | 2.95 | .89 | 1.08 – 5.00 | .15 (.36) | -.45 (.70) |  | 76 | 2.72 | .57 | 1.33 – 5.33 | .24 (.28) | .59 (.55) |
| Study 2 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| EPoC Drawing | 240 | 3.65 | 1.16 | 1.00 – 6.67 | .05 (.16) | .09 (.31) |  | 276 | 3.73 | 1.10 | 1.00 – 6.33 | -.07 (.15) | -.10 (.29) |
| Verbal Uncommon | 65 | .81 | .002 | .68 – .91 | -.17 (.30) | -.39 (.59) |  | 85 | .78 | .003 | .65 – .89 | -.22 (.26) | -.21 (.52) |
| Figural Uncommon | 38 | .84 | .002 | .76 – .92 | .23 (.38) | -.48 (.75) |  | 54 | .85 | .001 | .79 – .94 | .51 (.33) | .47 (.64) |
| Verbal Fluency | 65 | 13.23 | 25.99 | 4 – 25 | .32 (.30) | -.62 (.59) |  | 85 | 13.56 | 23.51 | 3 – 25 | .32 (.26) | -.31(.52) |
| Figural Fluency | 38 | 12.82 | 19.34 | 4 – 25 | .41 (.38) | .58 (.75) |  | 55 | 14.24 | 28.37 | 1 – 26 | .12 (.32) | -.09 (.63) |
|  | | | | | | | | | | | | | |
| *Note.* Standard errors in parentheses | | | | | | | | | | | | | |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 2**  *Male and female participants in regions of the distribution of storyboard creativity scores in Study 1* | | | | | | | |  |
|  |  | Male | | Female | |  |  |  |
| Distribution region | *N* | *n* | % | *n* | % | M/F ratio | *χ*2 (1) | *ϕ* |
| z < - 1 SD | 18 | 6 | 13.64 | 12 | 15.79 | .86 | .10 | .07 |
| -1 SD < z < 1 SD | 83 | 27 | 61.36 | 56 | 73.68 | .83 | 1.98 | .15 |
| z > 1 SD | 19 | 11 | 25.00 | 8 | 10.53 | 2.37 | 4.38\* | .48 |
|  |  |  |  |  |  |  |  |  |
| *Note.*% = percent within gender; M/F ratios are based on percent within gender to control for differences between the groups in sample size; *ϕ* = phi coefficient, which may be interpreted using Cohen's (1992) guidelines for *r* (﻿.10 = small, .30 = medium, and .50 = large).  *\*p* < .05 | | | | | | | | |

**Discussion**

Taken together, the results of Study 1 do not support the GMVH in creativity. Although Levene’s test may have been statistically significant with more participants or equal group sizes, the observed VR = 1.56 was smaller than commonly suggested criteria (VR ≤ 3.00; Dean & Voss, 1999; Keppel et al., 1992; Kirk, 2013) and those typically found in research comparing two groups (VR = 2.51; Ruscio & Roche, 2019). Men were significantly overrepresented in the upper tail of the distribution of scores. However, although the M/F ratios indicated a slightly greater proportion of women than men in the mid-range and lower tail of the distribution, the ratios were not significant and the effects were small. In sum, the pattern observed here does not reflect the GMVH (i.e., overrepresentation of men in both the lower and upper tails of the distribution). Additionally, men and women did not differ significantly in mean scores on the storyboard task.

**Study 2**

**Method**

***Participants***

This study was approved by the Institutional Review Board at (approval #redacted). Participants were adolescents recruited from middle and high schools in the Northeastern United States. Data for those not reporting either male or female gender were excluded from analyses, resulting in a sample of 529 participants (53.7% female, 46.3% male). Participants’ ages ranged from 11 to 18 years old (*M* = 13.28, *SD* = 1.70). Participants’ school level ranged from 7 to 12, with the majority of participants in grades 7 or 8 (79.4%). Race and Ethnicity were represented as follows: 65.8% Caucasian/White, 13.8% Asian, 3.6% Hispanic or Latinx, 2.1% African American/Black, 1.1% American Indian/Alaska Native, .9% Native Hawaiian or other Pacific Islander, and 1.7% other. Boys and girls did not differ significantly by age, *t* (523.29) = -.46, *p* = .64, *d* = .04, 95% CI [-.14, .21], school level, *χ*2 (5) = 8.08, *p* = .15, *ϕc =* .12, or racial and ethnic representation, *χ*2 (6) = 4.89, *p* = .56, *ϕc =* .10.

***Materials***

**Alternate Uses Task.** Verbal divergent thinking was assessed using an Alternate Uses Task (AUT; Guilford, 1956), in which participants were asked to list *“different, interesting and original ways… ideas that other children would not think of*” for a cardboard box. Participants were provided with a sheet of lined paper and given 8 minutes to write down as many ideas as possible. Responses were scored for uncommonness based on the probability for an individual to provide each response and averaged to create a total uncommonness score. The task was also scored for fluency (i.e., the total number of responses provided).

**Parallel Lines Task.** Figural divergent thinking was assessed using an adaptation of the repeated figures task (parallel lines) from the Torrance Tests of Creative Thinking-Figural (Torrance, 2008). Participants were provided with three sheets of paper containing a series of parallel lines and given eight minutesto complete as many different and original drawings as possible using a pair of lines. Participants were also asked to *“Write below each drawing what it is*.” Participants were given 8 minutes to complete the task. Responses were scored for uncommonness based on the probability for an individual to provide each response and averaged to create a total figural uncommonness score. The task was also scored for fluency (i.e., the total number of responses provided).

**EPoC Drawing Task.** Creative drawing was assessed using a graphic task from the Evaluation of Potential for Creativity (EPoC; Lubart et al., 2011). Participants were given one of three photo sheets (alternate forms) depicting 8 everyday objects and asked to invent an original drawing using four of the objects (in addition to any other elements they wanted to add). Participants were given 15 minutes to complete the drawing. Responses were rated for creativity using the CAT (Amabile, 1982) by three creativity researchers. Inter-rater reliability for scores, rated on a scale from 1 to 7 (with 7 being the most creative) was acceptable (α = .76). Scores did not differ significantly by test form, *F*(2, 513) = .66, *p* = .52. Total mean creativity scores were calculated by averaging ratings for each participant across raters.

***Procedure***

Data used in this study consists of the first measurement occasion of a longitudinal study on creativity and identity development in adolescence (redacted for blind review). Each measurement occasion consisted of an in-class, collective paper-and-pencil assessment (all performance-based tasks reported here), and an out-of-school assessment (self-report questionnaires not presented here, and demographic information). To address classic issues of repeated measurement in creativity assessment (Barbot, 2019), alternate forms of each performance-based task were distributed to different cohorts in a counterbalanced fashion across measurement occasions. Therefore, only a randomly selected subgroup of participants completed the cardboard AUT and parallel lines tasks during this first measurement occasion.

***Data Analyses***

All analyses conducted mirrored those used in Study 1.

**Results**

None of the score distributions by gender deviated substantially from normal, after removing the uncommonness score for one female outlier (*z* = -4.92) on the figural DT task. As seen in Table 1, there was a significant mean gender difference for uncommonness scores on the verbal DT task, *t*(148) = 2.55, *p* = .01, *d* = .42, 95% CI [.09, .75], with boys scoring higher than girls.[[5]](#footnote-6) No significant mean gender differences were found for the EPoC drawing task, *t*(514) = -.88, *p* = .36, *d* = -.08, 95% CI [-.25, .10], figural DT uncommonness, *t*(90) = -1.48, *p* = .14, *d* = -.31, 95% CI [-.73, .10],[[6]](#footnote-7) verbal DT fluency, *t*(148) = -.41, *p* = .68, *d* = -.07, 95% CI [-.39, .26], or figural DT fluency, *t*(91) = -1.36, *p* = .18, *d* = -.29, 95% CI [-.70, .13]. Levene’s test of equality of variance was not statistically significant for the EPoC drawing task, *F*(1, 514) = .01, *p* = .94, VR = 1.05, verbal DT uncommonness, *F*(1, 148) = .16, *p* = .69, VR = .67, figural DT uncommonness, *F*(1, 90) = 1.62, *p* = .21, VR = 2.00, verbal DT fluency, *F*(1, 148) = .28, *p* = .60, VR = 1.11, or figural DT fluency, *F*(1, 91) = 1.94, *p* = .17, VR = .68.[[7]](#footnote-8)

The total distribution of scores did not deviate from normal for verbal DT uncommonness (Skew = -.21, SE = .20; Kurtosis = -.25, SE = .39), figural DT uncommonness (Skew = -.27, SE = .25; Kurtosis = -.01, SE = .50), verbal DT fluency (Skew = .31, SE = .20; Kurtosis = -.48, SE = .39), figural DT fluency (Skew = .27, SE = .25; Kurtosis = .06, SE = .50), or EPoC drawing (Skew = -.02, SE = .11; Kurtosis = -.03, SE = .22). The M/F ratios within the regions of the distribution of scores demonstrated different patterns for the tasks (Table 3). For the distribution of scores on the verbal divergent thinking task, girls were significantly overrepresented in the lower tail (more than 1 SD below the mean) of the distribution, *χ*2 (1) = 5.16, p < .05, *ϕ* = .47. No other M/F ratios were significant, though the effect sizes for several of them indicated moderate effects (see Table 3).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 3**  *Male and female participants in regions of the distribution of divergent thinking and creativity scores in Study 2* | | | | | | | | |
|  |  | Male | | Female | |  |  |  |
| Distribution region | *N* | *n* | % | *n* | % | M/F ratio | *χ*2 (1) | *ϕ* |
| EPoC Drawing |  |  |  |  |  |  |  |  |
| z < - 1 SD | 62 | 31 | 12.92 | 31 | 11.23 | 1.15 | .35 | .08 |
| -1 SD < z < 1 SD | 382 | 178 | 74.17 | 204 | 73.91 | 1.00 | .00 | .00 |
| z > 1 SD | 72 | 31 | 12.92 | 41 | 14.86 | .87 | .40 | .08 |
|  |  |  |  |  |  |  |  |  |
| Verbal Uncommonness |  |  |  |  |  |  |  |  |
| z < -1 SD | 23 | 5 | 7.69 | 18 | 21.18 | .36 | 5.16\* | .47 |
| -1 SD < z < 1 SD | 104 | 46 | 70.77 | 58 | 68.24 | 1.04 | .11 | .03 |
| z > 1 SD | 23 | 14 | 21.54 | 9 | 10.59 | 2.03 | 3.40 | .38 |
|  |  |  |  |  |  |  |  |  |
| Figural Uncommonness |  |  |  |  |  |  |  |  |
| z < - 1 SD | 17 | 10 | 26.32 | 7 | 12.96 | 2.03 | 2.64 | .39 |
| -1 SD < z < 1 SD | 60 | 21 | 55.26 | 39 | 72.22 | .77 | 2.83 | .22 |
| z > 1 SD | 15 | 7 | 18.42 | 8 | 14.81 | 1.24 | .21 | .12 |
|  |  |  |  |  |  |  |  |  |
| Verbal Fluency |  |  |  |  |  |  |  |  |
| z < -1 SD | 23 | 11 | 16.92 | 12 | 14.12 | 1.20 | .22 | .10 |
| -1 SD < z < 1 SD | 123 | 53 | 81.54 | 70 | 82.35 | .99 | .02 | .01 |
| z > 1 SD | 4 | 1 | 1.54 | 3 | 3.53 | .44 | .56 | .37 |
|  |  |  |  |  |  |  |  |  |
| Figural Fluency |  |  |  |  |  |  |  |  |
| z < - 1 SD | 13 | 5 | 13.16 | 8 | 14.55 | .90 | .04 | .06 |
| -1 SD < z < 1 SD | 66 | 30 | 78.95 | 36 | 65.45 | 1.21 | 1.99 | .17 |
| z > 1 SD | 14 | 3 | 7.89 | 11 | 20.00 | .40 | 2.58 | .43 |
|  |  |  |  |  |  |  |  |  |
| *Note.*% = percent within gender; M/F ratios are based on percent within gender to control for differences between the groups in sample size,  *ϕ* = phi coefficient, which may be interpreted using Cohen's (1992) guidelines for *r* (﻿.10 = small, .30 = medium, and .50 = large).  *\*p* < .05 | | | | | | | | |

**Discussion**

Although results for several individual analyses in Study 2 may be interpreted as evidence for the GMVH, the overall pattern of results suggests that the GMVH is not generalizable across tasks and domains. Boys scored significantly higher than girls in uncommonness of the responses on the verbal divergent thinking task. No significant mean gender differences were found for any other task (though note that analyses for the figural DT task likely only had sufficient power to detect large effects). No significant gender differences in the variability of scores were found for any task. However, the VRs for the figural DT task suggest that these results may have been detected as significant with a larger sample size. The figural uncommonness VR = 2.00 indicates that boys had double the amount of variability in scores than girls. Although this number is still small compared to commonly suggested criteria (VR ≤ 3.00; Dean & Voss, 1999; Keppel et al., 1992; Kirk, 2013) and VRs typically found in research comparing two groups (VR = 2.51; Ruscio & Roche, 2019), it does exceed VRs interpreted as evidence for the GMVH in previous studies (e.g., He et al., 2013; He & Wong, 2011; Karwowski, Jankowska, Gajda, et al., 2016). However, the different directions of the VR for verbal creativity (suggesting greater variability for girls) and the VR for figural creativity (suggesting greater variability for boys) are noteworthy. The different directions of the VRs for verbal and figural divergent thinking uncommonness, negligible difference in variance for the creative drawing task, and different patterns of M/F ratios in different regions of the distributions are overall inconsistent with the basic tenets of the GMVH.

**General Discussion**

We examined gender differences in creative variability in verbal and figural creativity and divergent thinking in adults and adolescents. Overall, we found little evidence for the GMVH, which posits that men would exhibit greater variability in creativity scores and be overrepresented in the upper and lower extremes of the distribution of scores. In study one, the variability of scores on a creative writing task did not differ substantively between adult men and women. Although men were significantly overrepresented in the upper tail of the distribution of scores ( > 1 SD above the mean), men and women were similarly represented in both the mid-range and lower tails of the distribution. In study two, the variability of scores on figural and verbal divergent thinking and creative drawing tasks did not differ substantively between adolescent boys and girls (though see discussion for Study 2). However, the mean score for boys was significantly higher than that for girls in uncommonness of verbal divergent thinking. Most importantly, the pattern of the proportions of boys and girls in different regions of the distribution of scores differed by task. Boys and girls demonstrated fairly equivalent representation in different regions of the distribution of scores on the EPoC drawing. However, girls were significantly overrepresented in the lower tail of the distribution for verbal divergent thinking uncommonness, whereas boys were overrepresented (albeit not significantly) in the lower tail of the distribution of figural divergent thinking. Although these results do not match our expectation that males would demonstrate greater variability on figural (as opposed to verbal) divergent thinking tasks, it partially supports our assertion that gender differences in creative variability differ by domain and task.

**Theoretical Implications**

Our finding that gender differences in creative variability are inconsistent across domains and tasks suggests that the GMVH may not account for the greater prevalence of eminent creative men (as opposed to women). If greater male variability in creative ability explains men’s greater real-world creative achievement, this effect should be consistent across domains and tasks, as men have greater real-world creative achievement across domains (Baer & Kaufman, 2008; Piirto, 1991; Simonton, 1994). Although we did not find any statistically significant gender differences in the variability of creativity scores for any task in our adult or adolescent samples, inconclusive results for the figural DT task may have been due to a relatively low sample size. However, we also did not find evidence for gender differences in the variability of creativity scores that exceeded either (1) common criteria suggested for rejection of variance homogeneity (i.e., VR ≤ 3.00; Dean & Voss, 1999; Keppel et al., 1992; Kirk, 2013) or (2) the average VR for studies comparing two groups (i.e., VR = 2.51; Ruscio & Roche, 2019). We did find that men were significantly overrepresented in the upper tale of the distribution of scores for creative writing. Yet, creative writing has been suggested to be the most (relatively) equitable domain in terms of the composition of eminent male and female creators (Simonton, 1994). Thus, gender differences in creative achievement across domains may not be attributable to the GMVH.

Given that the GMVH has received support from several studies using the TCT-DP, it may be worth-while to examine exactly what the TCT-DP is picking up on in terms of gender differences. Our results for figural tasks in Study 2 are inconsistent with those reported by studies using the TCT-DP, suggesting that it may not be a simple matter of gender differences in figural versus verbal abilities. Our creative drawing task, rated for creativity using the CAT (widely considered to be the “gold standard” of creativity assessment; Baer & McKool, 2014; Cseh & Jeffries, 2019), and figural divergent thinking task did not provide support for greater male variability or for male overrepresentation in the upper tail of the distribution. The norming sample for the TCT-DP reveals that “A minimal trend in favour of boys is relatively consistent throughout the various sub-samples...” (Urban, 2005, p. 276), though no significant gender difference in mean scores was found for the total population. In a cross-cultural examination of the TCT-DP, boys’ mean scores were higher than that for girls in nine out of ten countries (Jellen & Urban, 1989). Although the difference was only statistically significant in the Philippines, many of the sample sizes, which ranged from *n* = 27 to 70 (*M* = 50), would not have been large enough to detect even a medium to large effect size with 80% power. Thus, the measurement equivalence of the TCT-DP across genders may be of concern when using the instrument to examine gender differences. In sum, the TCT-DP may suffer from a measurement bias that has not been considered and should be formally investigated, given that the GMVH is mainly grounded on findings relying on the TCT-DP.

Although developmental patterns of gender differences in creativity may indeed underly discrepancies in some of the results among studies using the TCT-DP (He, 2018), our results align with suggestions that additional boundary conditions of the GMVH exist (Ju et al., 2015; Karwowski, Jankowska, Gralewski, et al., 2016). Although there may be contexts in which men demonstrate greater creative variability and/or are overrepresented in the extremes of a distribution, our results reveal that there are contexts in which this is untrue. Identifying under what circumstance the GMVH is – and is not – supported will provide a more precise understanding of the causes and consequences of gender differences in creative variability.

Although both biological and socio-cultural factors have been theorized as explanations for gender differences in the variability of creativity (e.g., Karwowski et al., 2016), potential explanatory variables are rarely studied empirically. For instance, biological and neurocognitive differences between men and women (e.g., hormonal differences or differences in brain connectivity) have been suggested to underlie differences in cognition, but how these factors actually contribute to differences in creative ability or performance remains unclear (see Abraham, 2016)*.* Similarly, socio-cultural explanations, such as gender-based expectations regarding social roles and opportunities for developing certain skills, have been proposed, yet are rarely measured (though see Simonton, 1992). Researchers reporting that creativity variability varies across contexts (e.g., rural vs. urban samples; Ju et al., 2015) provide support for the influence of socio-cultural variables, but have thus far discussed creative performance as a generalized construct. The results in Study 2 show that the pattern of the distribution of creativity scores differs as a function of domains and/or task modalities *within the same sample*, suggesting that there are potential explanatory variables that most likely (1) reflect individual-level differences and (2) operate at the domain or task level. One potential variable that may vary with gender and lead to different outcomes by task is cognitive strategy, by which the cognitive strategies that are more-or-less used by males versus females may result in enhanced creativity on one task and diminished creativity on another (Abraham, 2016). However, this is speculative and, because the pattern of results found in the current study is not consistent with expectations or with previous work on the GMVH in other cognitive abilities (e.g., Feingold, 1992), it would be premature to speculate about the underlying cause of results for each task. More research is needed to determine variables that may contribute to the observed gender patterns of creativity variability, as they interact with the creativity task at hand.

**Limitations and Recommendations for Future Research**

There are several limitations of our study that may be addressed in future work. First, non-significant tests for the figural DT task in Study 2 should be interpreted with caution, as the number of participants who completed the task was not sufficient to detect small to medium effect sizes for differences in mean performance between genders. However, mean gender differences in performance were not the focus of the present study. Rather, the focus was on gender group differences in variability, which was notably operationalized by Levene’s test and VRs, as classically used in this line of work. This points to perhaps a larger issue, which cannot be resolved by our study: that of using Levene’s test and VRs to assess meaningful differences in variability between groups. Although implications for the GMVH are clearer when these analyses are supplemented by examining the distributions, exactly what constitutes a meaningful difference in variability needs to be resolved to ensure the consistent interpretation of results across studies examining the GMVH in creativity. Second, the creative outcomes assessed in our study were limited. Our creative writing task and creative drawing tasks correspond with the performance and arts domains (respectively) identified by Kaufman and colleagues (Kaufman, 2012; McKay et al., 2016). Although the divergent thinking tasks used in our study reflect differences in stimuli and response type (figural and verbal), how divergent thinking relates to creative performance in specific domains is uncertain (e.g., Runco & Acar, 2012). Future research should include a broader set of domains and tasks. Third, scores were based on a single task per indicator of creativity (e.g., verbal divergent thinking was assessed by one alternate uses task). Participants may perform differently on identical tasks when presented with different stimuli, such as parallel lines versus circles on a figural divergent thinking task (see Barbot, 2018; Barbot et al., 2016, 2020). Thus, future studies could examine creative variability in different domains by administering multiple tasks of the same indicators of creativity to the same participants. Finally, the suggestion that our results differ from that of studies using the TCT-DP due to our use of different indicators of creativity could be strengthened by including the TCT-DP in addition to other measures in future investigations. Because our study is the first to investigate the GMVH in creativity in a North American sample, the possibility that our results differ from previous studies due to differences in cultural background, geographical location, ethnicity, and/or language cannot be ruled out. However, differences arising from these contextual features would also support the instability of the GMVH in creativity.

Identifying and measuring potentially confounding variables in future studies would allow researchers to control for their influence on creative outcomes. Although using ANCOVA to detect mean differences between groups controlling for other variables is familiar to many researchers, group differences in variability controlling for these variables may also be examined by focusing on analysis of the residuals (see Rosopa et al., 2013; Schroeder et al., 2020). Propensity score matching (i.e., pairing members of each group based on numerous similar characteristics), may also be used in future studies to control for the potential confounding effects of other variables (see Austin, 2011). Matching male and female participants who are similar to one another on multiple relevant variables would help to establish if gender differences in creativity means or variability exist independent of the influence of these variables. Studies examining the GMVH in creativity, as well as many studies examining mean gender differences, have thus far not controlled for other variables and would benefit from using these methods.

**Conclusion**

The current study examined the GMVH, which posits that men exhibit greater variability in creativity scores and are overrepresented in the upper and lower extremes of the distribution of scores. In two studies, we found little evidence for gender differences in creative variability for verbal and figural task performance in adults and adolescents. Additionally, we found that the representation of males and females in different regions of the distribution of creativity scores differs across domains and tasks. In light of evidence that gender differences in creative variability are inconsistent across domains and tasks, broad claims about the causes and consequences of gender differences in creativity based on the GMVH should be avoided. Identifying the specific contexts in which gender differences in creative variability may be found will provide greater insight into why these differences occur, as well as the outcomes that can be attributed to them.

**References**

Abraham, A. (2016). Gender and creativity: An overview of psychological and neuroscientific literature. *Brain Imaging and Behavior*, *10*(2), 609–618. https://doi.org/10.1007/s11682-015-9410-8

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. (1996). *Creativity in context: Update to “The Social Psychology of Creativity.”* Westview Press.

Arden, R., & Plomin, R. (2006). Sex differences in variance of intelligence across childhood. *Personality and Individual Differences*, *41*(1), 39–48. https://doi.org/10.1016/j.paid.2005.11.027

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, *46*(3), 399–424. https://doi.org/10.1080/00273171.2011.568786

Baer, J. (1993). *Divergent thinking and creativity: A task-specific approach*. Lawrence Erlbaum Associates Publishers.

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. (2015). The Importance of Domain-Specific Expertise in Creativity. *Roeper Review*, *37*(3), 165–178. https://doi.org/10.1080/02783193.2015.1047480

Baer, J., & Kaufman, J. C. (2005). Bridging generality and specificity: The Amusement Park Theoretical (APT) model of creativity. *Roeper Review*, *27*, 158–163.

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

Baer, J., & McKool, S. S. (2014). The gold standard for assessing creativity. *International Journal of Quality Assurance in Engineering and Technology Education*, *3*(1), 81–93. https://doi.org/10.4018/ijqaete.2014010104

Barbot, B. (2018). The dynamics of creative ideation: Introducing a new assessment paradigm. *Frontiers in Psychology*, *9*, 1–8. https://doi.org/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/https://doi.org/10.1037/aca0000232

Barbot, B. (2020). Creativity and self-esteem in adolescence: A study of their domain-specific, multivariate relationships. *Journal of Creative Behavior*, *54*(2), 279–292. https://doi.org/10.1002/jocb.365

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

Carson, S. H., Peterson, J. B., & Higgins, D. M. (2005). Reliability, validity, and factor structure of the Creative Achievement Questionnaire. *Creativity Research Journal*, *14*(1), 37–50. https://doi.org/10.1207/s15326934crj1701

Cohen, J. (1992). A power primer. *Quantitative Methods in Psychology*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Conti, R., Coon, H., & Amabile, T. M. (1996). Evidence to support the componential model of creativity: Secondary analyses of three studies. *Creativity Research Journal*, *9*(4), 385–389.

Cseh, G. M., & Jeffries, K. K. (2019). A scattered CAT: A critical evaluation of the consensual assessment technique for creativity research. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(2), 159–166. https://doi.org/10.1037/aca0000220

Dean, V., & Voss, D. (1999). *Design and analysis ofexperiments*. Springer-Verlag.

Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use welch’s t-Test instead of student’s t-Test. *International Review of Social Psychology*, *30*(1), 92–101. https://doi.org/10.5334/irsp.82

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Feingold, A. (1992). Sex differences in variability in intellectual abilities: A new look at an old controversy. In *Review of Educational Research* (Vol. 62, Issue 1).

Guilford, J. P. (1956). The structure of intellect. *Psychological Bulletin*, *53*(4), 267–293. https://doi.org/10.1037/h0040755

He, W. (2018). A 4-year longitudinal study of the sex-creativity relationship in childhood, adolescence, and emerging adulthood: Findings of mean and variability analyses. *Frontiers in Psychology*, *9*, 1–14. https://doi.org/10.3389/fpsyg.2018.02331

He, W., & Wong, W. (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., Wong, W., 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*(5220), 41–45. https://doi.org/10.1126/science.7604277

Jellen, H. G., & Urban, K. K. (1989). Assessing creative potential world-wide: The first cross-cultural application of the Test for Creative Thinking—Drawing Production (TCT-DP). *Gifted Education International*, *6*(2), 78–86. https://doi.org/10.1177/026142948900600204

Jeon, K.-N., Moon, S. M., & French, B. (2011). Differential effects of divergent Thinking, domain knowledge, and interest on creative performance in art and math. *Creativity Research Journal*, *23*(1), 60–71. https://doi.org/10.1080/10400419.2011.545750

Johnson, W., Deary, I. J., & Carothers, A. (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., Gajda, A., Marczak, M., Groyecka, A., & Sorokowski, P. (2016). Greater male variability in creativity outside the WEIRD world. *Creativity Research Journal*, *28*(4), 467–470. https://doi.org/10.1080/10400419.2016.1229978

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

Kaufman, J. C. (2012). Counting the muses: Development of the Kaufman Domains of Creativity Scale (K-DOCS). *Psychology of Aesthetics, Creativity, and the Arts*, *6*(4), 298–308. https://doi.org/10.1037/a0029751

Kaufman, J. C., & Baer, J. (2004). Sure, I’m creative--but not in mathematics!: Self-reported creativity in diverse domains. *Empirical Studies of the Arts*, *22*(2), 143–155. https://doi.org/10.2190/26HQ-VHE8-GTLN-BJJM

Kaufman, J. C., Cole, J. C., & Baer, J. (2009). The construct of creativity: Structural model for self-reported creativity ratings. *Journal of Creative Behavior*, *43*(2), 119–134. https://doi.org/10.1002/j.2162-6057.2009.tb01310.x

Keppel, G., Saufley, W. H., & Tokunaga, H. (1992). *Introduction to design and analysis: A student’s handbook (2nd ed.)*. Freeman.

Kirk, R. E. (2013). *Experimental design. Procedures for the behavioral sciences (4th ed.)*. SAGE Publications.

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

Lubart, T. I., Besançon, M., & Barbot, B. (2011). *Evaluation of Potential for Creativity (EPoC)*. Hogrefe.

Machin, S., & Pekkarinen, T. (2008). Assessment: Global sex differences in test score variability. In *Science* (Vol. 322, Issue 5906, pp. 1331–1332). https://doi.org/10.1126/science.1162573

McKay, A. S., Karwowski, M., & Kaufman, J. C. (2016). Measuring the muses: Validating the Kaufman Domains of Creativity Scale (K-DOCS). *Psychology of Aesthetics, Creativity, and the Arts*. https://doi.org/10.1037/aca0000074

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

Piirto, J. (1991). Why are there so few? (Creative women: Visual artists, mathematicians, musicians). *Roeper Review*, *13*(3), 142–147.

Plucker, J. A. (1999). Reanalyses of student responses to creativity checklists: Evidence of content generality. *Journal of Creative Behavior*, *33*(2), 126–137. https://doi.org/10.1002/j.2162-6057.1999.tb01042.x

Plucker, J. A. (2004). Generalization of creativity across domains: Examination of the method effect hypothesis. *Journal of Creative Behavior*, *38*(1), 1–12. https://doi.org/10.1002/j.2162-6057.2004.tb01228.x

Plucker, J. A., Beghetto, R. A., & Dow, G. T. (2004). Why isn’t creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educational Psychologist*, *39*(2), 83–96. https://doi.org/10.1207/s15326985ep3902\_1

Qian, M., & Plucker, J. A. (2018). Looking for renaissance people: Examining domain specificity-generality of creativity using item response theory models. *Creativity Research Journal*, *30*(3), 241–248. https://doi.org/10.1080/10400419.2018.1488348

Rawlings, D., & Locarnini, A. (2007). Validating the Creativity Scale for Diverse Domains using groups of artists and scientists. *Empirical Studies of the Arts*, *25*(2), 163–172. https://doi.org/10.2190/q538-2307-2627-1256

Rosopa, P. J., Schaffer, M. M., & Schroeder, A. N. (2013). Managing heteroscedasticity in general linear models. *Psychological Methods*, *18*(3), 335–351. https://doi.org/10.1037/a0032553

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., 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. https://doi.org/10.1007/978-1-4419-1467-5

Ruscio, J., & Roche, B. (2012). Variance heterogeneity in published psychological research: A review and a new index. *Methodology*, *8*(1), 1–11. https://doi.org/10.1027/1614-2241/a000034

Ruscio, J., Whitney, D. M., & Amabile, T. M. (1998). Looking inside the fishbowl of creativity: verbal and behavioral predictors of creative performance. *Creativity Research Journal*, *11*(3), 243–263. https://doi.org/10.1207/s15326934crj1103\_4

Schroeder, A. N., Rosopa, P. J., Whitaker, J. H., Fairbanks, I. N., & Xoxakos, P. (2020). Heteroscedasticity in organizational research. In E. F. Stone-Romero & P. J. Rosopa (Eds.), *Research in Human Resource Management: Toward Valid Research-based Inferences* (pp. 67–86). Information Age.

Shadish, W., Cook, T., & Campbell, T. (2002). *Experiments and quasi-experimental designs for generalized causal inference*. Houghton-Mifflin. http://impact.cgiar.org/pdf/147.pdf

Silvia, P. J., Wigert, B., Reiter-Palmon, R., & Kaufman, J. C. (2012). Assessing creativity with self-report scales: A review and empirical evaluation. *Psychology of Aesthetics, Creativity, and the Arts*, *6*(1), 19–34. https://doi.org/10.1037/a0024071

Simonton, D. K. (1992). Gender and genius in japan: Feminine eminence in masculine culture. *Sex Roles*, *27*(3–4), 101–119. https://doi.org/10.1007/BF00290012

Simonton, D. K. (1994). *Greatness: Who makes history and why*. Guilford Press.

Taylor, C. L., & Kaufman, J. C. (2020). Values across creative domains. *Journal of Creative Behavior*, Advance online publication. https://doi.org/10.1002/jocb.470

Taylor, C. L., Kaufman, J. C., & Barbot, B. (2020). Measuring Creative Writing with the Storyboard Task: The Role of Effort and Story Length. *Journal of Creative Behavior*. https://doi.org/10.1002/jocb.467

Torrance, E. P. (2008). *The Torrance tests of creative thinking norms-technical manual figural (streamlined) forms A and B*. Scholastic Testing Service.

Urban, K. K. (2004). Assessing Creativity: The Test for Creative Thinking - Drawing Production (TCT-DP) The Concept, Application, Evaluation, and International Studies. *Psychology Science*, *46*(3), 387–397.

Urban, K. K. (2005). Assessing creativity: The Test for Creative Thinking - Drawing Production (TCT-DP). *International Education Journal*, *6*(2), 272–280.

Urban, K. K., & Jellen, H. G. (1996). *Test for Creative Thinking-Drawing Production (TCT-DP)*. Swets and Zeitlinger.

Vartanian, O., Smith, I., Lam, T. K., King, K., Lam, Q., & Beatty, E. L. (2020). The relationship between methods of scoring the alternate uses task and the neural correlates of divergent thinking: Evidence from voxel-based morphometry. *NeuroImage*, *223*, Advance online publication. https://doi.org/10.1016/j.neuroimage.2020.117325

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

﻿

1. As in any observational study, there are numerous factors that vary with gender that could explain differences in creativity variability (Shadish et al., 2002). Potential methods of controlling for individual differences in factors that may covary with gender are highlighted in the discussion. [↑](#footnote-ref-2)
2. The initial CAT ratings from one rater exhibited irregularities for task B2 and was subsequently re-scored by the same rater. [↑](#footnote-ref-3)
3. Although Levene’s test (based on the absolute value of the residuals around the mean)is used in the current studies, alternative methods of detecting differences in variability between groups may be used in future studies. For instance, the Brown-Forsythe test is preferable when distributions deviate from normal (Rosopa et al., 2013; Schroeder et al., 2020). [↑](#footnote-ref-4)
4. Given the current sample parameters, it is not surprising that mean gender differences in performance were not observed (besides it was not the main focus here). Indeed, a meta-analysis highlighted how negligible the typical mean gender differences in creativity are (Hedges’ *g* = .06 for overall creativity, and .04 for product-based assessments; Thompson, 2016), requiring very large sample sizes to detect a difference as significant. In the present study, a power sensitivity analysis showed that a *d* = .54 would need to be observed for the *t*-test to be statistically significant (using G\*Power 3.1; Faul, Erdfelder, Buchner, & Lang, 2009). [↑](#footnote-ref-5)
5. Because different methods of scoring the AUT can result in different relationships with other variables (e.g., Vartanian et al., 2020), we also conducted our analyses using an alternate method of scoring the originality of DT responses. No gender differences in mean scores or variability were found when bracket scoring was used (e.g., points assigned for responses provided by a certain percentage of respondents). [↑](#footnote-ref-6)
6. The number of participants who completed the figural DT task would only be sufficient for 80% power to detect a medium to large effect size for a two-tailed *t*-test (approx. *d* > .6; according to G\*Power 3.1; Faul, Erdfelder, Buchner, & Lang, 2009). Therefore, non-significant results for this test should be interpreted with caution. [↑](#footnote-ref-7)
7. Based on He's (2018) findings that differences in variability may emerge only after the age of 16, we ran identical analyses with participants separated in two age groups. Differences in variability were not detected on any task with participants under the age of 16 (*ns* ranging from 80 to 438). Differences in variability were also not detected in participants aged 16 or over for the creative drawing task (*n* = 77). The number of participants over the age of 16 who completed the divergent thinking tasks were not sufficient to examine variability (*n* ranging from 12 to 21). [↑](#footnote-ref-8)