EXAMINING GENDER DIFFERENCES IN CREATIVITY

Kuan Chen Tsai
University of the Incarnate World

Abstract
According to creativity literature, the results about gender differences in creativity are mixed. The main purpose of this study was to examine the gender differences in creative performance by using the Consensual Assessment Technique (CAT). The majority of creativity studies focus on children and young adults. The focus of the current study, however, was on older adults. The results of this study suggest that gender differences using the CAT to evaluate creativity in terms of collage do exist. Male adults excelled females in creative performance of collage making-tasks, however, from a technical perspective, there was no difference between the two groups. With regard to the educational level (graduate and doctoral), no difference was found in creativity and technical goodness. Taken as a whole, the CAT shows some value in examining creative performance in adult learners.

Keywords. Gender differences, creativity, the Consensual Assessment Technique

Introduction
Creativity literature suggests that individuals’ background traits have some influence on creativity (Ai, 1999; Batey & Furnham, 2006). Among the background characteristics, gender is one of the investigated targets (Charyton, 2006). According to the literature, the results are mixed (Barron & Harrington, 1981). Some studies revealed no gender difference between males and females (Baer & Kaufman, 2008; Charyton & Snelbecker, 2007), whereas others revealed that gender differences in creative performance do exist (Hoff, 2005; Matud, Rodriguez, & Grande, 2007).

An appropriate measure of creative performance has been an interesting and intriguing topic for students of creativity (Feldhusen & Goh, 1995; Houtz & Krug, 1995; Urban, 2005). According to a preceding discussion in the literature, it is clear that different tests paint different pictures of this phenomenon. In fact, based on the literature, it shows this inconsistency regarding which one outperforms another (Kaufman, Baer, Agars, & Loomis, 2010). In addition, several cross-cultural studies indicated that environmental and cultural factors might affect creative performance in terms of gender differences.

A popular approach is the Consensual Assessment Technique (CAT; Amabile, 1982, 1996), which is widely used and well validated in creativity research, especially in diverse experiment conditions (Baer, Kaufman, & Gentile, 2004). In the CAT, participants are asked to create poems, collages, or stories and those products will be independently evaluated by experts with the criteria of novelty and appropriateness. Experts judge artifacts based on their own perception of creativity. Generally, interrater reliabilities among judges are quite satisfactory, ranging from .70 to .90 (Amabile, 1996; Hennessey, 2003; Hennessey & Amabile, 1988).

The main idea of the CAT is based on the assumption that the best measure of creative artifacts comes from the collective judgment of experts in that
field (Kaufman, Plucker, & Baer, 2008). The procedure of CAT is similar to the evaluation of creative works or other kinds of accomplishments in the real world such as Nobel Prizes, Grammy Awards, and Pulitzer Prizes. As Kaufman, Baer, Cole, & Sexton, (2008) indicated, “for creativity assessment in the real world, it is common for panels of experts in a given domain to be asked to evaluate the creativity of some creative product or group of products” (p. 171). In addition, Kaufman et al. (2010) believed that, “The CAT is a powerful tool used by creativity researchers in which groups of expert judges rate the creativity of a set of creative products (such as stories, collages, poems, etc.)” (p. 201). One of the advantages of CAT is that its validity is not contingent on that of any particular creativity theory because its key tenet is requiring artifacts being evaluated by experts in the specific domain (Amabile, 1996; Baer, 1993). Kaufman, Bear, and Cole (2009) also suggested when using CAT for evaluating creativity, “it is probably safest to… use experts whenever possible” (p. 231).

The main purpose of this study was to examine the gender differences in creative performance by using the Consensual Assessment Technique. The majority of creativity studies focus on children and young adults. The focus of the current study, however, was on older adults. By doing so, some light will be shed on adult educators who are willing to encourage creativity in adult classrooms.

**Literature Review: Creativity and Gender**

Core characteristics associated with creativity—for example, age, gender, intelligence, and personality—have been heavily investigated in the last two decades (Barron & Harrington, 1981). The issues of gender differences in abilities and in assessment have also been discussed in the study of creativity (Baer & McKool, 2009; Kaufman et al., 2010). In fact, the creativity tests are designed to avoid gender-biased results (Kim, 2007; Torrance, 1974). A number of studies have shown no gender differences in creativity (Kaufman, 2006; Kogan, 1974). Baer and Kaufman (2008) provided a comprehensive review on the question of gender differences in creativity. In terms of scores on divergent thinking tests, they revealed the following: the evidence does not clearly support gender differences in creativity based on test results; however, to the extent that a case for such gender differences can be made, the available evidence suggests that women and girls tend to score higher on creativity tests than men and boys. (Baer & Kaufman, 2008, p. 78)

Overall, the lack of differences in genders is found in many creativity studies. Baer and Kaufman (2008) then argued that any gender differences in creativity probably stems from an environmental factor. They also believed in “relative equality in creative ability” between the genders (Baer & Kaufman, 2008, p. 76).

However, there are several studies that have revealed that a gender difference in creative performance was found. Kaufman (2006) examined self-report measures of creativity in 3,553 students across different domains. He found that females rated themselves higher on social-communications and on the visual-artistic factor, while males rated themselves higher on the science-analytic factor and sports. However, there were no differences on the verbal-artistic factor (p. 1074). These results are consistent with research on gender and self-reported intelligence and this self-assessment is also consistent with gender stereotypes.

Stolitzfus, Nibbelink, Vredenburg, and Thyrum (2011) used the Torrance Test of Creative Thinking (TTCT; Torrance, 1974) as an approach to evaluate creativity in undergraduate students. They discovered male students’ performance on creativity was better than that of females. He and Wong (2011) used the Test for Creative Thinking—Drawing
Production (TCT-DP; Urban & Jellen, 1996) to investigate gender differences in creativity among schoolchildren. Their results showed that boys outperformed girls in boundary-breaking thinking, whereas girls were better in thoroughness of thinking. They suggested both genders have their strengths and weaknesses in terms of different dimensions of creative thinking.

The discrepancy in creativity, in regard to gender, was also found in different cultures. Oral et al. (2007) utilized two different tests—the Alternate Uses Test (Christensen, Guilford, Merrifield, & Wilson, 1960) and the Consequences Test (Christensen, Merrifield, & Guilford, 1958)—to measure creativity of schoolchildren in Turkey. The results showed no gender differences on the Alternate Uses Test, but females scored higher on the Consequences Test. Cheung and Lau (2010) used the Wallach-Kogan Creativity Tests (Wallach & Kogan, 1965) on schoolchildren in Hong Kong. Their results showed middle school girls outperformed boys in several dimensions of creativity. By canonical correlation analysis, Ai (1999) found Spanish male and female students showed different aspects of creativity related to academic achievement. For male students, flexibility was the predominant factor that related to academic performance, while in females, fluency and elaboration were more important factors. He suggested that the gender differences in creativity could be in part influenced by gender roles.

Methods
Participants

The participants in the study consisted of 18 graduate or doctoral students from two classes in a private southwest university. The sample included 7 males and 11 females, with a mean age of 42.67 years (SD = 13.47 years). The demographic breakdown was as follows: 2 Asians, 2 African Americans, 3 Caucasians, 9 Hispanics, and 2 mixed backgrounds.

Procedure

Participants first read and signed a consent form, and then were given instructions for the task. Participants were given 20 minutes to complete a collage. During this activity, the experimenter gave the salient instruction, “Use your imagination and play with the material to create a collage.” The instruction was given to participants as follows:

You are invited to create a collage. You will be provided a set of pre-cut construction paper shapes in a variety of colors, a bottle of glue, and a blank white A3 paper. You will not be given a pair of scissors. Rather, you will need to tear the paper with your hands and use the glue to complete the collage. The reason is that we want you to play with the material and have fun. The topic of the collage is “As an adult, how do you perceive the learning?” Please use your “imagination and creativity” to finish this activity. You will have 20 minutes to create your unique collage. Before you start, please complete the background information you have received. Hope you enjoy this activity!

Rating Procedures

Three experts (two from the fashion department and one an art educator) were given the collages in different, randomly assigned order and asked to rate them for creativity, technical goodness, and aesthetic value on a 1 to 5 scale. To be consistent with the CAT methodology, the experts were asked to rate these three dimensions of the collages, working independently. They were not asked to explain or defend their ratings but to use their own personal sense of what is creative, technical, and aesthetic in the domain. The Instruction in the grading sheet was given to each expert:

There is no one criterion in rating these collages in terms of creativity, technical goodness, and aesthetic value. The topic of the collage is “As an adult, how do you
perceive the learning?” For the purpose of this study, the researcher will not any criteria for you; rather, you are asked to rate the sketches solely on the basis of your thoughtful-but-subjective opinions of their creative products. You are asked to rate the creativity, technical goodness, and aesthetic value for the collage on a 5-point rating scale from 1 (the lowest level of the dimension) to 5 (the highest level of the dimension). Please circle the number on the grading sheet. Thank you.

After judges graded each collage in the grading sheet, the inter-rater reliability was checked. It was estimated using Cronbach’s coefficient alphas (Cronk, 2008). Kaufman, Baer, Cole, and Sexton (2008) indicated, “Coefficient alpha is a standard measure of internal consistency, and has been used in creativity research as a measure of interrater reliability (treating raters as items)” (p. 174). The reliability test of Cronbach’s Alpha for creativity was .760, for technical goodness .754, and for aesthetic value -.157. This result showed that the consistency among the raters was sufficient only for creativity and technical goodness (Kaufman et al., 2008). As a result, the dimension of aesthetic value was dropped for further analysis.

Results

A Pearson correlation coefficient was calculated for the relationship between participants’ age, GPA, creativity, and technical goodness. As Table 1 shows, a moderate negative correlation was found between GPA and creativity, $r(16) = -0.627, p < .5$. A moderate negative correlation that was not significant was found between age and creativity, $r(16) = -0.432, p > .05$. The relationship between creativity and technical goodness showed a weak but not significant negative correlation.

<table>
<thead>
<tr>
<th>Variables</th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Age</td>
<td>42.67</td>
<td>13.47</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. GPA</td>
<td>3.87</td>
<td>.14</td>
<td>.322</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Creativity</td>
<td>3.31</td>
<td>.85</td>
<td>-.432</td>
<td>-.627*</td>
<td>--</td>
<td></td>
</tr>
</tbody>
</table>

* $p < .05$.

In evaluating the gender difference of creativity and technical goodness, an independent-samples $t$ test was calculated comparing the mean scores of the two groups. Table 2 displays that only creativity was significantly different between male and female, $t(16) = 2.758, p < .05$. The mean of the male group was significantly higher ($M = 3.95, SD = .911$) than the mean of the female group ($M = 2.91, SD = .518$). Large effects size was also found (Cohen’s $d = 1.403$). In terms of technical goodness, no significant difference was found, $t(16) = -0.407, p > .05$. In addition, an independent-samples $t$ test was calculated comparing the mean scores of creativity and technical goodness between graduate and doctoral students. As Table 3 shows, no significant difference was found either in creativity, $t(16) = 1.89, p > .05$, or in technical goodness, $t(16) = .097, p > .05$. 


### Table 2

*Gender Difference for Creativity and Technical Goodness (n = 18)*

<table>
<thead>
<tr>
<th>Variables</th>
<th>Male</th>
<th>Female</th>
<th>t (16)</th>
<th>p</th>
<th>Cohen’s d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creativity</td>
<td>3.95</td>
<td>2.91</td>
<td>2.758</td>
<td>.023</td>
<td>1.403</td>
</tr>
<tr>
<td>Technical goodn</td>
<td>3.14</td>
<td>3.33</td>
<td>-.407</td>
<td>.691</td>
<td>-.198</td>
</tr>
</tbody>
</table>

### Table 3

*Educational Level Difference for Creativity and Technical Goodness (n = 18)*

<table>
<thead>
<tr>
<th>Variables</th>
<th>Graduate</th>
<th>Doctoral</th>
<th>t (16)</th>
<th>p</th>
<th>Cohen’s d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creativity</td>
<td>3.76</td>
<td>3.03</td>
<td>1.89</td>
<td>.081</td>
<td>.918</td>
</tr>
<tr>
<td>Technical Goodn</td>
<td>3.29</td>
<td>3.24</td>
<td>.097</td>
<td>.924</td>
<td>.053</td>
</tr>
</tbody>
</table>

In order to further understanding the relationship among age, GPA, creativity, and technical goodness, a simple linear regression was calculated. As table 4 shows, only GPA could predict creativity. A significant regression equation was found, $F(1, 9) = 5.82, p = .039$, with an adjust $R^2 = .325$, indicating medium effect sizes (Cohen, 1988). Participants’ GPA was a significant predictor of creativity. Interestingly, the higher the GPA, the lower creative score will be attained.

### Table 4

*Linear Regression Analysis for GPA Predicting Creativity*

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>19.00</td>
<td>6.535</td>
<td>2.907</td>
<td>.017</td>
<td></td>
</tr>
<tr>
<td>GPA</td>
<td>-4.08</td>
<td>1.690</td>
<td>-.627</td>
<td>-2.413</td>
<td>.039</td>
</tr>
</tbody>
</table>

*Note. $R^2 = .393$ (N = 18).*

### Discussion

Before discussing of the major findings of this study, several limitations should be considered. First, because of available samples, this study did not use randomized samplings. The sample size was also small. It is expected that recruiting more subjects would increase the power of the study. Second, this study utilized the collage-making activity as a criterion to evaluate creativity. There are other types of creative performance, such creative poem or story writing. In other words, if the participants had experienced different activities, the results may have been different. Finally, this study utilized the CAT as the assessment of creativity. It is possible that by using other evaluation methods might lead to different results.

The results of this study suggest that gender differences using the CAT to evaluate creativity in terms of collage do exist. Male adults excelled females in creative performance of collage making-tasks, however, from a technical perspective, there was no difference between the two groups. With regard to educational level (graduate and doctoral), no difference was found in creativity and technical
goodness. It is possible that males have a tendency to demonstrate boundary-breaking thinking (He & Wong, 2011), so that when they created a collage, the structure of pictures was more abstract and original. As a result, judges evaluated those collages with higher creative scores. It should be noted that initially three dimensions were included in the evaluation process: creativity, technical goodness, and aesthetics. However, the perspective of aesthetics did not attain an agreement among the three judges. It is possible that because two of them were from the fashion field and the third was from arts. Thus, they hold different views of aesthetics.

Another major finding shows a negative correlation between creativity and academic performance. In addition, the results of a linear regression show that the higher GPA, the lower creativity scores. This suggests that (a) creativity and intelligence are different constructs and (b) academic performance might affect creativity. This observation is in line with creativity literature (Amabile, 1996; Simonton, 1999). For example, Mednick and Andrews (1967) investigated a relationship between creativity and IQ, by using the Remote Associates Test (RAT; Medicak, 1962) as a creative thinking tool and the Scholastic Aptitude Test of verbal and mathematical ability (SAT-V & SAT-M) as indicators of IQ. The results from scores of high school students and college freshmen showed the opposite tendency between creativity and intelligence. The higher the scores on SAT, the lower the scores on RAT. As a result, they concluded intelligence and creativity are relatively independent.

Taken as a whole, the CAT shows some value in examining creative performance in adult learners. The current study reveals that gender differences in creativity do exist. To some extent academic performance seems to become a roadblock for creativity. Several limitations to this study should be recognized. First, the current study used only a small sample. Future researchers could recruit more participants to further understand gender differences in creativity. Second, the collage task served as an indicator of creative performance. However, it should be noticed that different kinds of creative tasks might lead to a different picture. In fact, the collage-making activity may show only a particular facet of creativity. Therefore, researchers could consider alternative assessments of creativity such as story writing, poem writing, or drawing activities. Finally, the expertise of raters might influence the evaluation of creativity. The current study used judges from two different fields, fashion design and fine arts. It is possible to yield different results from judges who are in the same field or judges from other suitable fields such as art educators.

References


Handbook of research on assessment technologies, methods, and applications in higher education (pp. 65-77). Hershey, PA: IGI Global.


