HHS Public Access

Author manuscript

Think Skills Creat. Author manuscript; available in PMC 2023 March 02.

Published in final edited form as:

Think Skills Creat. 2023 March ; 47: . doi:10.1016/j.tsc.2023.101240.

Evaluative Skill in the Creative Process: A Cross-cultural Study

Yawei Guo,
BNU-HKBU United International College

Shengjie Lin,
Yale University

Zachary J. Williams,
Vanderbilt University

Yuntao Zeng,
University of Macau

Lili Q. Cole Clark
University of Georgia

Abstract

Evaluative skill is the ability to accurately judge ideas on creativity (or originality), which is a critical component of creativity. Various aspects of creativity have been examined cross-culturally, but little research has focused on evaluative skill. The first goal of this study was to examine the measurement invariance of evaluative skill assessments, which were based on two types of divergent thinking tests (Line Meanings and Uses), between American (n = 341) and Chinese (n = 345) college students. Multi-group confirmatory factor analyses supported a two-factor model based on two types of evaluation tasks, and this model satisfied configural and weak invariance. However, partial strong invariance was satisfied only for the Uses evaluation task. Based on this evidence, our second goal was to explore the differences in evaluative skill between these two groups. Via latent mean comparisons, we found that American participants had better performance on evaluative skill based on the Uses evaluation task than their Chinese counterparts. Taken

Correspondence concerning this article should be addressed to Shengjie Lin, Center for Emotional Intelligence, Child Study Center, Yale University, 350 George Street, New Haven, CT 06511, shengjie.lin@yale.edu.
Yawei Guo, Faculty of Science & Technology, BNU-HKBU United International College; Shengjie Lin, Center for Emotional Intelligence, Child Study Center, Yale University; Zachary J. Williams, School of Medicine, Vanderbilt University; Yuntao Zeng, Faculty of Education, University of Macau; Lili Q. Cole Clark, Franklin College of Arts and Sciences, University of Georgia.

Author Statement

YG: Conceptualization, investigation, methodology, software, formal analysis, writing-original draft, writing-review & editing, validation, visualization
SL: Conceptualization, validation, writing-review & editing
ZW: Methodology, formal analysis, writing-review & editing
YZ: Conceptualization, writing-review & editing
LC: Investigation, writing-review & editing

Disclosure Statement: The author declares that we have no relevant or material financial interests that relate to the research described in this paper.

together, this study is one of the first to examine the cross-cultural differences in evaluative skill between American and Chinese adults. This study offered preliminary results showing some invariance in evaluative skill assessments across cultures and indicating cross-cultural differences in this ability.

Keywords

creativity; evaluative skill; cross-cultural; measurement invariance

1. Introduction

J. K. Rowling's bestseller *Harry Potter* was initially rejected by many publishers and the flying machine invented by the Wright Brothers was ignored by the US military at first (Licuanan et al., 2007). In addition to these eminent creative contributions (Big-C and Pro-C; Kaufman & Beghetto, 2009), creativity on lower levels (little-c and mini-c) may also be discounted (Benedek et al., 2016; Karwowski, 2007). For example, in workplaces, supervisors may underestimate employees' creative ideas and suggestions that could have otherwise contributed to the company's innovation. These examples illustrate that if people evaluate a novel idea as common or unoriginal, they may lose the opportunity to act on it, thereby missing out on any potential benefits. On the other hand, if individuals judge a commonplace idea as creative, they may misallocate scarce resources to that idea, resulting in expensive failures or the inability to support ideas with better outcomes (Cropley, 2006).

These concerns regard people's *evaluative skill* (Runco & Smith, 1992), defined as the ability to accurately identify creative ideas. This ability has been included in most creativity theories (e.g., Campbell, 1960; Cropley, 2006; Reiter-Palmon et al., 2012; Simonton, 1988; Wallas, 1926) and demonstrated to be highly relevant to creative achievement and creative idea implementation (Kasof, 1995; Runco & Chand, 1994; Runco & Smith, 1992). However, some scholars have noted that the research on the evaluative component of creativity is not as developed as one might expect (Kozbelt, 2007; Silvia, 2008), particularly when compared to other aspects of creativity like idea generation (or divergent thinking). There are two possible reasons for this. The first is that one may simply regard evaluative thinking as critical thinking or convergent thinking. This may be a misconception because there is psychometric evidence from previous studies demonstrating the discriminant validity of those evaluative measures (Runco, 1991; Runco & Dow, 2004; Runco & Smith, 1992). Specifically, these studies failed to find significant correlations between convergent thinking or critical thinking and evaluative accuracy. Second, the complexity of measuring evaluative skill may be responsible for the lack of studies in the creativity literature. To capture evaluative skill, there should be a list of ideas produced first (Runco, 2020), then participants are required to evaluate these ideas, and finally an evaluative accuracy index needs to be obtained from the raw scores of respondents (Benedek et al., 2016; Silvia, 2008), which may complicate the design of the study. Therefore, more research is warranted to examine evaluative skill.

The neglect of evaluative skill is also evident in its lacking in cross-cultural examinations. Cultural context contributes to different perceptions and manifestations of creativity (Glăveanu et al., 2020; Niu & Kaufman, 2013). In the past decade, researchers have compared aspects of creativity across cultures and societies (e.g., Glăveanu et al., 2019; Guo et al., 2021; Katz-Buonincontro et al., 2021; Karwowski et al., 2020; Niu & Kaufman, 2013; Zhang et al., 2021). Most of the effort has been spent comparing divergent thinking among individuals from different societies (e.g., Americans vs. Chinese; Europeans vs. Asians; Guo et al., 2021; Niu & Sternberg, 2002). However, we barely know whether individuals from different cultural backgrounds display different levels of evaluative skill. This may be related to the above-mentioned complexity in assessing evaluative skill. This evaluative ability varies widely based on social, cultural, and historical backgrounds (Glăveanu et al., 2019; Ivancovsky et al., 2019). Thus, it is difficult to know whether the measure of evaluative skill is reliable and valid cross-culturally (see Glăveanu et al., 2019 for more details about measuring creativity across cultures). In addressing this challenging question, testing of measurement invariance has become an emerging approach in examining whether an aspect of creativity has the same meaning when interpreted by individuals from different cultural groups (e.g., Guo et al., 2021; Katz-Buonincontro et al., 2021; Kim et al., 2006). This is a critical step because cross-cultural comparisons on creativity can only be meaningful and interpretable if the measurement invariance of the assessment is satisfied across cultures (Byrne & Watkins, 2003; Guo et al., 2021). Therefore, the first goal of the current study was to examine whether the measurement invariance of evaluative skill assessments can be established across groups from two cultures (i.e., American and Chinese college students). Second, if the measurement invariance was satisfied, we aimed to explore the differences in evaluative skill between individuals from these two cultural groups.

1.1. Evaluative Skill

1.1.1. Definition and measures—Creativity is generally regarded as the production of both novel and appropriate ideas or products (Plucker et al., 2004; Runco & Jaeger, 2012). Creativity researchers have made remarkable progress in helping people understand aspects of creativity like idea generation. However, far less is known about another important component of creativity: idea evaluation (Basadur, 1995; Silvia, 2008; Runco & Chand, 1994), a critical step with important implications for refining, monitoring, and implementing creative ideas (Licuanan et al., 2007). Indeed, the importance of idea evaluation is extensively recognized in most models of creativity (Mumford et al., 1991; Reiter-Palmon et al., 2012; Simonton, 1988; Wallas, 1926). It is suggested that idea evaluation plays a positive role in creative performance (Cropley, 2006; Runco & Chand, 1994). For instance, Lubart (1994) found that evaluation promotes more creative performance in story-writing and drawing tasks among college students. In addition, an emerging body of literature has demonstrated that evaluative skill is related to divergent thinking (Benedek et al., 2016; Guo et al., 2019; Stemler et al., 2020), one of the most widely studied creative components. Particularly, a recent meta-analysis (Guo et al., 2022) revealed a modest positive relationship between these two constructs.

1.1.2. Cross-cultural studies of creativity—Different cultures have different values and social norms, which may have a profound influence on various aspects of creativity

(Glăveanu et al., 2019; Kim, 2016; Lubart, 2010; Niu & Kaufman, 2013; Niu & Sternberg, 2002). Most cross-cultural comparisons of creative processes have been conducted between respondents from individualistic cultures such as the US and collectivistic cultures such as Japan and China (Niu & Kaufman, 2013; Puente-Diaz et al., 2016). In general, this body of literature tends to report higher creativity in Western compared to Eastern individuals (see Ng, 2001 and Shao et al., 2019 for a review). For example, Zha et al. (2006) showed that American graduate students performed better on divergent thinking tests than those from China. Ivancovsky and colleagues (2019) found that Israeli students had higher performance on the Alternative Uses task than their Korean and Japanese peers. However, there is also opposing evidence. For example, some studies found that East Asians had higher scores on divergent thinking tests than their Western counterparts (Cheung et al., 2016; Saad et al., 2015). As noted in these examples, thus far, most of the cross-cultural creativity studies have centered on the idea generation phase of creativity (e.g., divergent thinking), which leaves the idea evaluation stage (e.g., evaluative skill) understudied. Creativity contains multiple facets; therefore, cross-cultural findings on idea generation do not necessarily extend to idea evaluation (Ivcevic, 2022; Reiter-Palmon & Schoenbeck, 2019). Therefore, more cross-cultural research on other aspects of creativity is warranted.

1.1.3. Evaluative skill in Western and Eastern cultures—We barely know whether there are cross-cultural differences in evaluative skill. However, some insights from the creativity and cross-cultural psychology literature may shed light on this topic. Researchers have argued that culture can exert an influence on the evaluation criteria of creativity (Karwowski, 2016; Niu & Sternberg, 2006). The two fundamental criteria of creativity are the originality and appropriateness of ideas or products (Runco & Jaeger, 2012). When it comes to the valuation of these two aspects, it is suggested that Westerners think highly of originality, whereas their Eastern counterparts emphasize appropriateness or usefulness (Adair & Xiong, 2018; Morris & Leung, 2010; Nijstad et al., 2010; Xie & Paik, 2019). This divergence in priorities can potentially be explained by the different philosophical roots and cultural orientations of the two cultures (Niu & Sternberg, 2006). Western culture tends to be more individualistic and regards creativity as a form of divine inspiration; thus, people brought up in this culture may believe that creativity refers to things which are groundbreaking. In contrast, Eastern culture is more collectivistic and mainly based on Confucianism, which emphasizes family, benevolence, and hierarchical relationships (Chen & Chung, 1994; Hui & Triandis, 1986; Kim, 2009; Triandis, 2001). To them, new things which are created should be morally good and adhere to tradition; thus, their focus is on appropriateness. Taken together, it is possible that there are cross-cultural differences in evaluative skill among individuals from Western and Eastern cultural backgrounds.

To our knowledge, only one study has empirically investigated the cross-cultural differences in idea evaluation, among Japanese, Korean, and Israeli samples (Ivancovsky et al., 2019). This study found that individuals from East Asian backgrounds evaluated ideas more stringently than their Western counterparts due to more restrictive cultural norms. Although this study made a novel contribution to this area, it focused on evaluative stringency rather than evaluative accuracy. Specifically, this study investigated individuals' tendency to adopt lenient or strict criteria in judgment (i.e., the tendency to consider ideas as more or less

original, appropriate, useful, and deviant), but the cross-cultural differences in the accuracy of these creativity judgments were left unexamined. As mentioned in the introduction, both overestimating and underestimating creativity can be harmful to individuals and society; thus, accuracy in evaluation is critically important (Runco & Smith, 1992; Silvia, 2008). Therefore, more research which aims to examine the cultural differences in evaluative skill is needed.

1.2. Measurement Invariance

To answer the research question about whether there are differences in aspects of creativity between individuals from different cultures, researchers have suggested that measurement invariance, that the administered assessment measures the same underlying construct, should be ensured (Karwowski, 2016; Runco, 2004). In cross-cultural research, the test of measurement invariance is not only a powerful tool in examining the generalizability of an assessment across different populations, but also a critical pre-requisite for valid and meaningful cross-cultural comparisons (Byrne & Watkins, 2003; Jeong & Lee, 2019; Milfont & Fischer, 2010). In our case of evaluative skill, if the assumptions of measurement invariance are met, it would be valid for researchers to attribute the higher observed scores of one group to a higher level of evaluative skill as a latent trait. On the other hand, if the testing of measurement invariance fails, the group differences in observed scores across cultures may reflect biased measurements rather than true differences in the latent variable of interest (Jeong & Lee, 2019; Wicherts et al., 2005). To test measurement invariance, there are four steps: configural invariance, weak (also called metric) invariance, strong (also called scalar) invariance, and strict invariance (see Putnick & Bornstein, 2016 and Van de Schoot et al., 2012 for review). These four levels of invariance reflect (a) equivalent factor structures, (b) equivalent factor loadings, (c) equivalent item intercepts, and (d) equivalent residual (error) variances, respectively. In addition, researchers have also suggested partial measurement invariance (Cheung & Rensvold, 2002; Vandenberg & Lnce, 2000). That is, if some but not all factor loadings are the same, partial weak invariance can be considered met, and if some but not all intercepts are the same, partial strong invariance can be viewed as satisfied. Based on Thompson and Green (2006), at least partial strong MI should be established for meaningful latent mean comparisons.

In creativity research, although there is a growing interest in examining the measurement invariance of creativity assessments across gender, grade, and academic major (Miroshnik et al., 2022; Krumm et al., 2014; Krumm et al., 2016; Said-Metwaly et al., 2020), only a few studies have explored the measurement invariance of creativity assessments across cultures (Guo et al., 2021; Katz-Buonincontro et al., 2021; Kornilov et al., 2016). In Katz-Buonincontro and colleagues' study (2021), the measurement invariance of measures on self-beliefs about teaching for creativity between American and Chinese educators was investigated and the results supported partial measurement invariance in assessments of creative self-efficacy, creative mindset (both fixed and growth), and desirability of creativity for teaching success. However, Guo and colleagues (2021) found a more complex picture of measurement invariance based on two types of divergent thinking tests (Line Meanings and Real-world Problems) between US and Chinese college students. They found that

although weak invariance was met for fluency and originality on both tests, partial strong measurement invariance was only supported for fluency in the Real-world Problems.

1.3. The present study

The first goal of the present study was to examine the measurement invariance of the evaluative measures based on divergent thinking tests between American and Chinese college students. That is, we were interested in whether the latent structure of the evaluative skill assessments would stay invariant between individuals from these two cultures. The purpose of this goal was to ensure that the assessments are unbiased and meaningful in both populations. Second, if the measurement invariance was satisfied, we aimed to compare the latent means on evaluative skill between American and Chinese college students. The purpose of this aim was to better understand whether there are cross-cultural differences in evaluative skill, a critical component of creativity. Given that both research questions are exploratory, we did not have any specific hypotheses.

Methods

Participants

Data were collected in a large public university in the southeastern United States and a large public university in the central part of China. In total, there were 341 college students (243 females, 71.3%; M_{age} = 20.65, SD = 2.51) in the American group and 345 college students (288 females, 83.5%; M_{age} = 20.25, SD = 2.66) in the Chinese group. The ethnicities of the American sample included 61.9% White, 19.4% Asian, 7% Latino/a, 6.5% Native Hawaiian or Pacific Islander, 4.1% multiracial or other race/ethnicity, and 1.2% African American. Individuals from both groups had a wide range of majors, including art, science, and social science. All participants took part in an online survey containing the below-described measures.

Measures

To measure evaluative skill, there are two stages (Grohman et al., 2006; Runco, 2020; also see Benedek et al., 2016 and Silvia, 2008 for more details): In Stage One, ideas (usually from divergent thinking tests) are generated, and they are scored in terms of originality, which then serves as the criterion afterwards. In Stage Two, respondents are required to rate the ideas presented in terms of originality (Grohman et al., 2006; Runco, 2020; Runco & Smith, 1992; Runco & Vega, 1990), and these ratings are compared to the objective criterion obtained in Stage One. In the present study, in Stage One, ideas to be rated were collected from previous studies using divergent thinking tests of Line Meanings (to think of what each drawing could be: a curve and point figure; a wave-like figure) and Uses (to think of uses for each daily object: an umbrella; a book) among American and Chinese groups (Guo & Guo, 2021; Guo et al., 2021; over 300 participants from each cultural group). Please see Wallach and Kogan (1965) for more details about the Line Meanings and Uses tests. To illustrate, for the Line Meanings divergent thinking test, respondents in previous studies were presented with two figures (a curve and point figure and a wave-like figure) and asked to think of all the things each figure could represent. In the Line Meanings evaluation task in the present

study, for the curve and point figure, some examples of ideas that were rated include “wifi,” “sun,” and “music.”

All ideas to be rated in the current study were scored objectively in terms of originality (i.e., the percentage of people who proposed the idea out of the whole sample), which served as the criterion afterwards for calculating the evaluative accuracy index. It should be noted that originality refers to uncommonness here, although originality may have different meanings and methods of operationalization (Wilson et al., 1953). For idea compiling for evaluative measures, following Runco and colleagues’ method (Charles & Runco, 2001; Runco & Vega, 1990), based on response statistical frequency in previous studies, two highly original ideas (i.e., given by less than 5% of the sample), two moderately original ideas (i.e., given by more than 5% but less than 15% of the sample), and two common ideas (i.e., given by more than 15% of the sample) were selected from the idea pool for each of the subtests from each cultural group. All ideas selected met the additional requirement of being relevant to the task or appropriate. In other words, random or rare ideas that were irrelevant to the task were disregarded, and thus, highly original ideas here can be considered creative according to the standard definition of creativity (Plucker et al., 2004; Runco & Jaeger, 2012). Suggested by previous research (Ivancovsky et al., 2019), the six ideas selected may vary depending on the cultural group, although some were the same. In total, there were 24 ideas for evaluation and within each subtest the six ideas were presented in random order. For all 24 ideas, participants were asked to rate the ideas on a 1–10 scale with the following prompt: “*How many people out of 10 do you think can give this idea?*” Similar task instructions have been used by previous researchers (Charles & Runco, 2001; Grohman et al., 2006). Notably, asking participants to rate the infrequency of answers is a recommended practice because it is easier for people to operationalize compared to asking them to rate the originality (Charles & Runco, 2001).

Evaluative skill was operationalized as the rating accuracy. To obtain the accuracy index of originality ratings, the difference between participants’ ratings and the criterion was obtained (Grohman et al, 2006). Specifically, the accuracy index was calculated by subtracting the originality criterion (the percentage of people who proposed the idea out of the whole sample in previous studies) from respondents’ ratings (percentage value; for example, if a respondent answered that 4 out of 10 people can think of the idea, the value was 40%) on each presented idea to obtain an absolute value, which represented the discrepancy of judgement from the correct value. In this sense, higher values of discrepancy indicate lower levels of evaluative accuracy.

Data Analysis Strategy

The software *Mplus* version 8.4 (Muthén & Muthén, 2017) was used for all the confirmatory factor analyses (CFA). In general, to test whether MI at each step was established, the free baseline (or bottom-up) approach was applied (Stark et al., 2006); that is, a group of nested models with different levels of cross-group equality constraints gradually added were fitted and compared by means of examining the model $\chi 2$ change. First, to test configural invariance, we examined whether the evaluative measures had the same factors and relationships between indicators and factors across the two cultural groups. This was

tested by freeing all parameters across the two groups, which was the unconstrained baseline model. For each indicator-factor relationship, if the factor loading estimated across the two groups showed the same sign (e.g., both positive, both negative, or both zero) in addition to a good model fit, it suggested the satisfaction of configural invariance. Second, when configural MI was met, weak and strong MI were tested step by step. Specifically, to test weak invariance, we examined whether the factors were linked with corresponding indicators to the same extent (strength) across the two cultural groups; to test strong invariance, we examined whether American respondents obtained the same observed score on an indicator as their Chinese peers with the same ability on a respective latent factor. In other words, based on the unconstrained model, different levels of cross-group equality constraints were gradually added, including factor loadings and intercepts. If at a certain step, the more constrained model turned out to be significantly worse, then the current model was not acceptable and would be discarded, indicating the unsatisfaction of the hypothesized level of invariance. It should be cautioned that examining the model $\chi 2$ change to determine MI may be overly sensitive with a large sample size, leading to the rejection of MI despite trivial levels of variance between the groups. To solve this problem, other criteria have been suggested, for example, a change of −.01 in the comparative fit index (CFI; Cheung & Rensvold, 2002; Xu & Barnes, 2011), supplemented by a change of .015 in the root mean square error of approximation (RMSEA; Chen, 2007; Chen & West, 2008; Miroshnik et al., 2022). In other words, when compared to the less restricted model, if the ΔCFI value (reduced CFI) was less than or equal to .01, and the ΔRMSEA value (increased RMSEA) was less than or equal to .015, the more constrained model would be retained since it was more parsimonious. Since the second goal of the present study was to investigate differences in constructs of interest across groups via latent mean comparisons, that is, whether there are differences between the two cultural groups in evaluative skill at the latent level (i.e., as a latent trait), strong or partial strong MI needed to be met (Thompson & Green, 2006). Therefore, when strong invariance was violated, the other option of partial strong MI was tested. Potential partial strong models were tested based on the information provided by modification indices.

Results

Correlations Between Variables

For idea evaluation based on each of the four divergent thinking subtests, items were parceled for further analysis. Specifically, for the two subtests based on Line Meanings (12 ideas were evaluated in total), the average evaluative scores were obtained for the four highly original ideas, the four common ideas, and the four moderately original ideas respectively, resulting in three parcels (labelled as LME_ORI, LME_COM, LME_MOD in Table 1). Similarly, three parcels were obtained for the two subtests based on Uses (labelled as USEE_ORI, USEE_COM, USEE_MOD in Table 1). Therefore, for each respondent there were six evaluative scores that were analyzed further. For the American group, the internal reliability coefficients (coefficient alpha; Cronbach, 1951) across the three parcels based on Line Meanings were .761, and those across the three parcels based on Uses were .826. Those values were .776 and .843 respectively for the Chinese group. Table 1 displays means, standard deviations, and correlations between variables.

Confirmatory Factor Analyses

To examine the latent structure of the measures, CFA models with different numbers of factors were tested. Figure 1 displays the theoretical models to be tested. The one-factor model assumed the four subtests measured a universal construct of evaluative skill, whereas the two-factor model assumed distinctiveness of each kind of evaluative measure based on different types of divergent thinking tests (Runco & Dow, 2004). Table 2 shows the chi square statistic and fit indices for the two potential factor models when factor loadings were freely estimated in each cultural group as in the configural invariance model. There was a difference in chi square but not tested since they were not nested models. In terms of model fit, all indices suggested a significantly better fit of the two-factor model (CFI = .984, RMSEA = .071, SRMR = 0.029) than the one-factor model (CFI = .891, RMSEA = .171, SRMR = .063). The results supported that the six evaluative scores provide information about two components of evaluative skill. Thus, the following steps of testing measurement invariance were based on the two-factor model.

Testing of Measurement Invariance

Table 3 showed the parameter estimates based on the two-factor configural invariance model, and Table 4 displayed the model comparison results for each step of the measurement invariance tests. For the configural invariance model, with the factor loading on the first item set at 1 and those on the second item freely estimated, results showed positive factor loadings estimated for each indicator-factor relationship in each cultural group. To facilitate interpretation, standard coefficients were provided in Table 3. These estimates were significant; combined with the fit indices information indicating a good model fit (e.g., CFI = .984), these results together suggested the *configural invariance was supported* in the present analysis. In addition, the difference in CFI between the configural invariance model and the weak invariance model was .000, meeting the $\Delta CFI \leq .01$ criterion (and change in RMSEA $-.009 \leq .015$ criterion) and supporting *the assumption of weak invariance*. Last, the difference in CFI between the weak invariance model and the strong invariance model was .075, failing to meet the $\Delta CFI \leq .01$ criterion (and change in RMSEA $.074 > .015$ criterion) and suggesting *the violation of strong invariance*.

Next, potential partial strong models were tested. Examination of modification indices (M. I.) based on the strong invariance model indicated that parameter constraints on the intercepts of LME_ORI (M.I. = 102.278) and LME_MOD (M.I. = 41.619) could greatly affect model fit. Thus, a partial strong invariance model (with intercepts of LME_ORI and LME_MOD freely estimated between groups, i.e., testing the invariance of the Uses evaluation task) was fitted, and results revealed that when compared to the weak invariance model, the current partial strong model displayed a decrease in model fit, and the change in CFI was less than .01 (and the change in RMSEA was less than .015), suggesting that this model met the assumption of partial strong invariance.

Further, based on this partial strong invariance model, with the latent means in the U.S. group set at zero and those in the Chinese group freely estimated, the latent mean comparison showed higher evaluative scores of the Chinese sample on the Uses evaluation task compared to the American sample (Cohen's $d = 0.448$, $p < .001$). Since this evaluative

score refers to the discrepancy between respondents' ratings and the criterion values (i.e., lower values indicate high evaluative skill), this result indicated that the evaluative skill of the Chinese sample was lower than that of the American sample on the evaluation task based on Uses.

Discussion

As its first goal, the present study sought to examine the measurement invariance of the evaluative measures based on two different divergent thinking tests (Line Meanings and Uses) across American and Chinese adults. The second goal was to explore the differences in evaluative skill between individuals from these two groups. Overall, this study demonstrated that the Uses evaluation measures were sound tools for comparing individuals' evaluative skill between American and Chinese samples whereas the Line Meanings evaluation measures were not. Also, the findings indicated different levels of evaluative skill between Americans and Chinese when the evaluation task was based on Uses. In the following section, we discuss measurement invariance of the evaluative measures in cross-cultural contexts and the cross-cultural differences in evaluative skill between American and Chinese adults. We then discuss limitations and future directions, as well as the implications of the present study.

Measurement Invariance of Evaluative Measures

First, for the factorial structure of the measures, the two-factor model showed a better fit for both groups on evaluative scores. This finding is consistent with previous views that evaluation tasks based on different types of divergent thinking tests may not be interchangeable. For instance, Runco and Dow (2004) showed that participants had worse evaluative skill on the evaluation task based on Consequences compared to that based on Uses and Pattern Meanings.

In addition, in terms of measurement invariance, based on the two-factor model, different levels of measurement invariance tests with multi-group CFA analyses showed satisfied configural and weak invariance but violated strong invariance. First, the assumption of configural invariance indicated that the same two-factor latent structure for evaluative skill was assessed across American and Chinese groups. In other words, the measures have the same factors and relationships between indicators and factors across the two cultural groups. Second, the satisfaction of weak invariance indicated that researchers could compare the relationship between evaluative skill and other constructs of interest between American and Chinese college students. For instance, based on this evidence, studies can investigate whether the strength of the relationship between evaluative skill and creative self-efficacy is similar among American and Chinese adults. Third, the two-factor model did not meet the assumption of strong invariance, which indicates that the intercepts were not the same between the groups. That is, for the evaluative measures, American respondents do not obtain the same observed score on an indicator as their Chinese peers with the same ability on a respective latent factor.

The findings from testing the partial strong model demonstrated the performance of specific measures with regard to measurement invariance cross-culturally. Specifically, the partial

strong model (with intercepts of LME_ORI and LME_MOD freely estimated between groups, i.e., testing the invariance of the Uses evaluation task) showed a satisfying fit, indicating that the evaluation tasks based on Line Meanings are culturally dependent, and may be less applicable and valid in cross-cultural contexts. It is possible that Line Meanings (presenting figural stimuli) are interpreted differently by people from different cultures due to varying social and linguistic backgrounds (Niu & Sternberg, 2002). This is also consistent with a recent study (Guo et al., 2021) which found that Line Meanings failed to satisfy partial strong invariance between American and Chinese samples. Our study extends the literature by showing that the test of Line Meanings may not be valid tools in cross-cultural comparisons, not only as divergent thinking assessments, but also when ideas produced for Line Meanings are used as measures for evaluative skill.

Cross-cultural Differences in Evaluative Skill

For cross-cultural differences in evaluative skill, the findings from the latent mean comparison based on the partial strong model (with intercepts of LME_ORI and LME_MOD freely estimated between groups, i.e., testing the invariance of the Uses evaluation task) showed higher performance of American respondents on the Uses evaluation task compared to their Chinese counterparts. To our knowledge, our study is one of the first to investigate evaluative skill (accuracy) cross-culturally, and shows that American and Chinese adults perform differently in judging creative ideas, especially those produced for Uses. It should be noted that in the present study, evaluative skill was operationalized as individuals' ability to accurately identify uncommon ideas, rather than to judge the quality or appropriateness of ideas, which is another important aspect of idea evaluation beyond the scope of this study (see the limitations section for more details).

One explanation for the discrepancy is the different cultural orientations (individualism vs. collectivism) of the two groups, a classic argument in interpreting cross-cultural variations in creativity research (Lubart, 2010; Niu & Kaufman, 2013; Niu & Sternberg, 2006). Specifically, distinct cultural contexts may influence individuals' evaluation process. Embedded in a more individualistic context, Americans generally seek to differentiate themselves from the crowd and expect creative ideas or products to be novel or even groundbreaking (Morris & Peng, 1994; Xie & Paik, 2019). In contrast, coming from a more collectivistic background that values interdependence and harmonious relationships, Chinese individuals may expect a creative idea or product to be aligned with traditional and social norms (Niu & Kaufman, 2013). In addition, researchers have suggested that Chinese individuals prioritize the pragmatic aspects of creativity over novelty (Lan & Kaufman, 2012; Pang & Plucker, 2012). Taken together, with emphasis on appropriateness or usefulness, Chinese participants tend to value the originality aspect of creative ideas less than their American counterparts, and this focus on aspects other than originality may explain Chinese participants' lower evaluative skill based on Uses. Moreover, it is interesting to note that the Uses test (i.e., thinking of different uses for a daily object) itself may closely tap the usefulness focus in Chinese individuals' evaluation process; thus, when evaluating ideas produced for this type of divergent thinking test, the tendency to put less emphasis on the originality of ideas may become more salient. This is one potential reason for why cross-cultural differences were observed on the Uses evaluation task. Future studies should

examine whether East Asians have higher evaluative skill in judging the appropriateness or usefulness aspect of creative ideas compared to their Western counterparts.

Another possibility is that individuals in Chinese culture tend to take a holistic approach to the evaluation process (Choi & Nisbett, 2000; Nisbett et al., 2001). For example, traditional Chinese medicine takes a holistic approach to health that attempts to bring body, mind, and spirit into harmony (Sun et al., 2013; Wang et al., 2011). In the field of creativity, researchers have suggested that East Asians may employ more criteria in evaluating ideas than their Western counterparts (Lowenstein & Mueller, 2016). Accordingly, the lower performance on evaluating ideas based on Uses tasks might be a consequence of Chinese raters using broader criteria or having higher expectations in judging creative ideas. This echoes a recent study which found that Eastern (i.e., Japanese and Korean) participants rated ideas based on Uses more stringently than their Western (i.e., Israeli) counterparts (Ivancovsky et al., 2019). In other words, when evaluating ideas generated for Uses, Easterners tend to be stricter and give lower ratings of creativity compared to their Western counterparts.

Overall, this study extends the existing literature in several important ways. First, in terms of the components of creativity, our results showed a different picture than the previous cross-cultural creativity studies which mainly focused on the popularly researched divergent thinking. That is, this line of research tends to report higher creative potential (especially divergent thinking) in Western compared to Eastern individuals (see Ng, 2001; Shao et al., 2019 for a review). Our study extends the existing literature by showing differences between Americans and Chinese on evaluative skill (which was operationalized as the ability to identify uncommon ideas), a barely explored creativity construct in cross-cultural creativity research. This suggests that when referring to cross-cultural creativity differences, researchers should be specific about what type of creativity process or construct the research is interested in. Otherwise, they will produce an overly generalized misconception. Second, methodologically, evaluative skill was approached as a latent construct, and latent mean comparison was conducted to examine the cross-cultural differences to reach more statistically robust conclusions. Most previous studies have investigated the construct of evaluative skill at the observed level (which may indicate measurement error to some extent), as shown in the effect sizes analyzed in the latest meta-analysis (Guo et al., 2022). Furthermore, previous research found a cross-cultural difference in *group* idea selection (Erez & Nouri, 2010; Li et al., 2013). This indicates that in addition to a clear description of a specific creativity construct, cross-cultural research in this field should also consider the context (individual vs. group) in which the task is performed. More studies are needed to further explore whether cultural influences on evaluative skill vary depending on context.

Limitations and Future Directions

There are several limitations in this study. First, in terms of measures, the present study only employed six ideas presented for each evaluation task. Future studies should investigate evaluation tasks with more ideas to examine whether there are differences in the psychometric properties of the measures cross-culturally. In addition, although our study demonstrated that evaluative measures based on Uses are sound tools for assessing

evaluative skill among both American and Chinese participants, future studies should employ other types of divergent thinking tests (e.g., Instances and Pattern Meanings) as evaluation tasks. Other than evaluating ideas produced for divergent thinking tests, future studies should examine whether individuals from different cultural backgrounds evaluate creative behaviors (e.g., creative performance and creative products) differently. Furthermore, future researchers should examine individuals' evaluative skill on ideas from different sources. For example, rather than evaluating others' ideas like in the present study, future studies can explore cross-cultural differences in situations where people are evaluating their own ideas.

The second limitation concerns the sample. Specifically, the present study only focused on American and Chinese students who made it into a four-year college. Therefore, the findings may not apply to populations with different educational backgrounds or individuals in other societies. In future studies that examine cross-cultural differences in evaluative skill, researchers should include respondents of other educational backgrounds (e.g., adults who don't attend college) and go beyond the commonly used West-East comparison. Emerging research has shown that even within the often-regarded individualistic society, Americans from different states have varying perceptions of enforced rules and tolerance for deviance (tightness vs. looseness; Harrington & Gelfand, 2014). Therefore, the way different processes of creativity are operationalized in various cultures should be further addressed. Moreover, specific cultural values and personalities should be measured to provide explanations for potential cross-cultural differences in creativity.

Last, the present study operationalized evaluative skill as the ability to accurately identify uncommon ideas (which was popularly used in most studies analyzed in the meta-analysis by Guo et al., 2022), rather than the ability to judge the quality or appropriateness of ideas, which is another important aspect of idea evaluation (Benedek et al., 2016; Charles & Runco, 2001). Studies in the field of idea evaluation on appropriateness are scarce (e.g., Benedek et al., 2016; Charles & Runco, 2001; Guo et al., 2019; Runco & Charles, 1993). As mentioned in the discussion section, future studies are needed for investigating cross-cultural differences in that aspect, for example, examining whether Eastern respondents have higher performance in judging the quality or usefulness aspect of ideas compared to their Western counterparts.

Conclusion and Implications

Evaluative skill is an essential but rarely examined component of creativity in a cross-cultural context. The present study is one of the first to investigate the differences in evaluative skill across Eastern and Western cultures via statistically robust latent mean comparisons. As the first goal of the study, we tested the measurement invariance of evaluative skill assessments across two cultural groups. Specifically, it provides evidence that the measures of evaluative skill based on certain evaluation tasks failed to meet the assumptions of measurement invariance. With the rapid expansion of the cross-cultural creativity line of research, an increasing number of researchers have pointed out that cross-cultural comparison results may be biased if the measurement invariance of the assessments was not met (Guo et al., 2021; Karwowski, 2016). Our study provides an example of how

to examine whether instruments are psychometrically valid for assessing evaluative skill in a cross-cultural context. Accordingly, it is part of the collective effort to establish a set of measures that is appropriately invariant across multiple cultures.

As the second goal of the study, our findings contribute to the understanding that there are cross-cultural differences in people's evaluative skill. Specifically, it provides evidence that Westerners tend to perform better on evaluative skill based on certain evaluation tasks in addition to the popularly researched divergent thinking. Theoretically, it contributes to the cross-cultural creativity research, which has mostly focused on the idea generation phase, by extending studies in this area to the idea evaluation phase. Practically, given that evaluative skill is highly relevant to creative achievement and creative idea implementation (Kasof, 1995; Runco & Smith, 1992), educators should raise their awareness of training students' evaluative skill in addition to divergent thinking, particularly students from collectivistic cultures. For example, educators should provide more opportunities for students to practice their skills in evaluating others' ideas.

Acknowledgments

This article is based on the dissertation work of Yawei Guo. This work was partly supported by the Star-up Research Fund (UICR0700054-23) awarded to Yawei Guo from the BNU-HKBU United International College.

References

Adair WL, & Xiong TX (2018). How Chinese and Caucasian Canadians conceptualize creativity: the mediating role of uncertainty avoidance. Journal of Cross-Cultural Psychology, 49(2), 223–238. 10.1177/0022022117713153

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

Basadur M (1995). Optimal ideation-evaluation ratios. Creativity Research Journal, 8, 63–76.

Benedek M, Nordtvedt N, Jauk E, Koschmieder C, Pretsch J, Krammer G, & Neubauer AC (2016). Assessment of creativity evaluation skills: A psychometric investigation in prospective teachers. Thinking Skills and Creativity, 21, 75–84. 10.1016/j.tsc.2016.05.007

Brockner J, Ackerman G, Greenberg J, Gelfand MJ, Francesco AM, Chen ZX, … & Shapiro D (2001). Culture and procedural justice: The influence of power distance on reactions to voice. Journal of Experimental Social Psychology, 37(4), 300–315. 10.1006/jesp.2000.1451

Byrne BM, Shavelson RJ, & Muthén B (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. Psychological Bulletin, 105(3), 456. 10.1037/0033-2909.105.3.456.

Byrne BM, & Watkins D (2003). The issue of measurement invariance revisited. Journal of Cross-cultural Psychology, 34(2), 155–175. 10.1177/0022022102250225

Campbell DN (1960). Blind variation and selective retention in creative thought as in other knowledge processes. Psychological Bulletin, 67, 380–400.

Charles RE, & Runco MA (2001). Developmental Trends in the Evaluative and Divergent Thinking of Children. Creativity Research Journal, 13(3/4), 417–437. 10.1207/S15326934CRJ1334_19

Chen G-M, & Chung J (1994). The impact of Confucianism on organizational communication. Communication Quarterly, 42(2), 93–105. 10.1080/01463379409369919

Chen FF (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. Structural Equation Modeling: A Multidisciplinary Journal, 14(3), 464–504. 10.1080/10705510701301834

Chen FF, & West SG (2008). Measuring individualism and collectivism: The importance of considering differential components, reference groups, and measurement invariance. Journal of Research in Personality, 42(2), 259–294. 10.1016/j.jrp.2007.05.006

Cheung GW, & Rensvold RB (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. Structural Equation Modeling, 9(2), 233–255. 10.1207/S15328007SEM0902_5

Cheung PC, Lau S, Lubart T, Chu DHW, & Storme M (2016). Creative potential of Chinese children in Hong Kong and French children in Paris: A cross-cultural comparison of divergent and convergent-integrative thinking. Thinking Skills and Creativity, 22, 201–211. 10.1016/j.tsc.2016.09.005

Choi I, & Nisbett RE (2000). Cultural psychology of surprise: Holistic theories and recognition of contradiction. Journal of Personality and Social Psychology, 79(6), 890–905. https://psycnet.apa.org/doi/10.1037/0022-3514.79.6.890 [PubMed: 11138759]

Cordon SL, & Finney SJ (2008). Measurement invariance of the Mindful Attention Awareness Scale across adult attachment style. Measurement and Evaluation in Counseling and Development, 40(4), 228–245. 10.1080/07481756.2008.11909817

Cronbach LJ (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297–334. 10.1007/BF02310555

Cropley A (2006). In praise of convergent thinking. Creativity Research Journal, 18(2), 391–404. 10.1207/s15326934crj1803_13

Erez M, & Nouri R (2010). Creativity: The influence of cultural, social, and work contexts. Management and Organization Review, 6(3), 351–370.

Fee A, & Gray SJ (2012). The expatriate-creativity hypothesis: A longitudinal field test. Human Relations, 65 (12), 1515–1538. 10.1177/0018726712454900

Finch WH, & French BF (2015). Latent variable modeling with R. New York, NY: Routledge.

Garcia-Barrera MA, Kamphaus RW, & Bandalos D (2011). Theoretical and statistical derivation of a screener for the behavioral assessment of executive functions in children. Psychological Assessment, 23(1), 64. https://psycnet.apa.org/doi/10.1037/a0021097 [PubMed: 21244172]

Glăveanu VP (2019). Measuring creativity across cultures: Epistemological and methodological considerations. Psychology of Aesthetics, Creativity, and the Arts, 13(2), 227–232. 10.1037/aca0000216

Glăveanu VP, Hanchett Hanson M, Baer J, Barbot B, Clapp EP, Corazza GE, … & Sternberg RJ (2020). Advancing creativity theory and research: A socio- cultural manifesto. The Journal of Creative Behavior, 54(3), 741–745.

Grohman M, Wodniecka Z, & Klusak M (2006). Divergent thinking and evaluation skills: Do they always go together? Journal of Creative Behavior, 40(2), 125–145. 10.1002/j.2162-6057.2006.tb01269.x

Guo J, Ge Y, & Pang W (2019). The underlying cognitive mechanisms of the rater effect in creativity assessment: The mediating role of perceived semantic distance. Thinking Skills and Creativity, 33, 100572. 10.1016/j.tsc.2019.100572

Guo J, & Guo Y(2021). The influence of culture on creativity and its underlying motivational mechanisms. Unpublished manuscript.

Guo J, Lin S, & Guo Y (2018). Sex, Birth Order, and Creativity in the Context of China's One-Child Policy and Son Preference. Creativity Research Journal, 30(4), 361–369. 10.1080/10400419.2018.1530535

Guo Y, Lin S, Acar S, Jin S, Feng Y, Xu X, Zeng Y (2022). Divergent thinking and evaluative skill: A meta-analysis. Journal of Creative Behavior, 0(0), 1–17. 10.1002/jocb.539

Guo Y, Lin S, Guo J, Lu ZL, & Shangguan C (2021). Cross-cultural measurement invariance of divergent thinking measures. Thinking Skills and Creativity, 41, 100852. 10.1016/j.tsc.2021.100852

Harrington JR, & Gelfand MJ (2014). Tightness–looseness across the 50 united states. Proceedings of the National Academy of Sciences, 111(22), 7990–7995. 10.1073/pnas.1317937111

Ivancovsky T, Shamay-Tsoory S, Lee J, Morio H, & Kurman J (2019). A dual process model of generation and evaluation: A theoretical framework to examine cross-cultural differences in the creative process. Personality and Individual Differences, 139, 60–68. 10.1016/j.paid.2018.11.012

Jeong S, & Lee Y (2019). Consequences of not conducting measurement invariance tests in cross-cultural studies: A review of current research practices and recommendations. Advances in Developing Human Resources, 21(4), 466–483. 10.1177/1523422319870726

Katz- Buonincontro J, Hass R, Kettler T, Tang LM, & Hu W (2021). Partial measurement invariance of beliefs about teaching for creativity across US and Chinese educators. British Journal of Educational Psychology, 91(2), 563–583. [PubMed: 33180977]

Karwowski M (2007). Teachers' nominations of students' creativity: Should we believe them? Are the nominations valid? The Social Sciences, 2, 264–269.

Karwowski M (2016). Culture and psychometric studies of creativity. In Glăveanu VP, (Ed.), The Palgrave handbook of creativity and culture research (pp. 159–186). Palgrave Macmillan.

Karwowski M, Gralewski I, Patston T, Cropley DH, & Kaufman JC (2020). The creative student in the eyes of a teacher: A cross-cultural study. Thinking Skills and Creativity, 35, Article 100636. 10.1016/j.tsc.2020.100636

Kasof J (1995). Explaining creativity: The attributional perspective. Creativity Research Journal, 8, 325–370. 10.1207/s1326934crj0804_1

Kaufman JC, & Beghetto RA (2009). Beyond big and little: The four c model of creativity. Review of General Psychology, 13(1), 1–12. 10.1037/a0013688

Kenny DA, Kaniskan B, & McCoach DB (2015). The performance of RMSEA in models with small degrees of freedom. Sociological Methods & Research, 44{3), 486–507. 10.1177/0049124114543236

Kim D (2016). Role of cognitive flexibility in bilingualism and creativity (Doctoral dissertation, University of Georgia, Athens, USA).

Kim KH (2009). Cultural influence on creativity: The relationship between Asian culture (Confucianism) and creativity among Korean educators. The Journal of Creative Behavior, 43(2), 73–93. 10.1002/j.2162-6057.2009.tb01307.x

Kim KH, Cramond B, & Bandalos DL (2006). The latent structure and measurement invariance of scores on the Torrance Tests of Creative Thinking–Figural. Educational and Psychological Measurement, 66(3), 459–477. 10.1177/0013164405282456

Kornilov SA, Kornilova TV, & Grigorenko EL (2016). The cross- cultural invariance of creative cognition: A case study of creative writing in US and Russian college students. New Directions for Child and Adolescent Development, 2016(151), 47–59.

Kozbelt A (2007). A quantitative analysis of Beethoven as self-critic: Implications for psychological theories of musical creativity. Psychology of Music, 35(1), 144–168. 10.1177/0305735607068892

Krumm G, Filipppetti VA, Lemos V, Koval J, & Balabanian C (2016). Construct validity and factorial invariance across sex of the Torrance Test of Creative Thinking–Figural Form A in Spanish-speaking children. Thinking Skills and Creativity, 22, 180–189. 10.1016/j.tsc.2016.10.003

Krumm G, Lemos V, & Filippetti VA (2014). Factor structure of the Torrance Tests of Creative Thinking Figural Form B in Spanish-speaking children: Measurement invariance across gender. Creativity Research Journal, 26(1), 72–81. 10.1080/10400419.2013.843908

Kuhn JT, & Holling H (2009). Measurement invariance of divergent thinking across gender, age, and school forms. European Journal of Psychological Assessment, 25(1), 1–7. 10.1027/1015-5759.25.1.1

Li C, Kwan L, Liou S, & Chiu C-Y (2013). Culture, group processes, and creativity. In Yuki M, & Brewer M (Eds.). Culture and group processes (pp. 145–165). New York, NY: Oxford University Press.

Licuanan BF, Dailey LR, & Mumford MD (2007). Idea evaluation: Error in evaluating highly original ideas. The Journal of Creative Behavior, 41(1), 1–27. 10.1002/j.2162-6057.2007.tb01279.x

Lin S, Falbo T, Qu W, Wang Y, & Feng X (2021). Chinese only children and loneliness Stereotypes and realities. American Journal of Orthopsychiatry, 91(4), 531–544. 10.1037/ort0000554 [PubMed: 34166053]

Lonergan DC, Scott GM, & Mumford MD (2004). Evaluative aspects of creative thought: Effects of appraisal and revision standards. Creativity Research Journal, 16, 231–246. 10.1080/10400419.2004.9651455

Lowenstein J, & Mueller J (2016). Implicit theories of creative ideas: How culture guides creativity assessments. Academy of Management Discoveries, 2(4), 320–348. 10.5465/amd.2014.0147

Lubart TI (1994). Product centered self-evaluation and the creative process. Unpublished doctoral dissertation, Yale University, New Haven, CT.

Lubart T (2010). Cross-cultural perspectives on creativity. The Cambridge handbook of creativity, 265–278.

Lucas HC, & Goh JM (2009). Disruptive technology: How Kodak missed the digital photography revolution. The Journal of Strategic Information Systems, 18(1), 46–55. 10.1016/j.jsis.2009.01.002

Meade AW, & Lautenschlager GJ (2004). A comparison of item response theory and confirmatory factor analytic methodologies for establishing measurement equivalence/invariance. Organizational Research Methods, 7(4), 361–388. 10.1177/1094428104268027

Milfont TL, & Fischer R (2010). Testing measurement invariance across groups: Applications in cross-cultural research. International Journal of psychological research, 3(1), 111–130.

Miroshnik KG, Shcherbakova OV, & Kaufman JC (2022). Kaufman Domains of Creativity Scale: Relationship to Occupation and Measurement Invariance across Gender. Creativity Research Journal, 34(2), 159–177. 10.1080/10400419.2021.1953823

Morris MW, & Leung K (2010). Creativity East and West: Perspectives and parallels. Management and Organization Review, 6, 313–327.

Mueller J, Melwani S, Loewenstein J, & Deal JJ (2018). Reframing the decision-makers' dilemma: Towards a social context model of creative idea recognition. Academy of Management Journal, 61(1), 94–110. 10.5465/amj.2013.0887

Mumford MD (2001). Something old, something new: Revisiting Guilford's conception of creative problem solving. Creativity Research Journal, 13, 267–276. 10.1207/S15326934CRJ1334_04

Mumford MD, Mobley MI, Uhlman CE, Reiter-Palmon R, & Doares L (1991). Process analytic models of creative capacities. Creativity Research Journal, 4, 91–122.

Muthén LK & Muthén BO (1998-2017). Mplus user's guide (8th ed.). Muthén & Muthén.

Ng AK (2001). Why Asians are less creative than westerners. Singapore: Prentice Hall.

Nijstad BA, De Dreu CK, Rietzschel EF, and 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. 10.1080/10463281003765323

Nisbett RE, Peng K, Choi I, & Norenzayan A (2001). Culture and systems of thought: holistic versus analytic cognition. Psychological Review, 108(2), 291–310. https://psycnet.apa.org/doi/10.1037/0033-295X.108.2.291 [PubMed: 11381831]

Niu W, & Kaufman JC (2013). Creativity of Chinese and American cultures: A synthetic analysis. The Journal of Creative Behavior, 47(1), 77–87. 10.1002/jocb.25

Niu W, & Sternberg RJ (2002). Contemporary studies on the concept of creativity: The East and the West. The Journal of Creative Behavior, 36(4), 269–288. 10.1002/j.2162-6057.2002.tb01069.x

Niu W, & Sternberg RJ (2006). The philosophical roots of Western and Eastern conceptions of creativity. Journal of Theoretical and Philosophical Psychology, 26(1-2), 18–38. https://psycnet.apa.org/doi/10.1037/h0091265

Plucker JA, Beghetto RA, & Dow GT (2004). Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. Educational Psychologist, 39(2), 83–96. 10.1207/s15326985ep3902_1

Puente-Díaz R, Cavazos-Arroyo J, & Vargas-Barrera F (2021). Metacognitive feelings as a source of information in the evaluation and selection of creative ideas. Thinking Skills and Creativity, 39, 100767. 10.1016/j.tsc.2020.100767

Puente-Diaz R, Maier MA, Brem A, & Cavazos-Arroyo J (2016). Generalizability of the four C model of creativity: A cross-cultural examination of creative perception. Psychology of Aesthetics, Creativity, and the Arts, 10(1), 14. https://psycnet.apa.org/doi/10.1037/aca0000038

Putnick DL, & Bornstein MH (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. Developmental review, 41, 71–90. [PubMed: 27942093]

Reiter-Palmon R, Wigert B, & de Vreede T (2012). Team creativity and innovation: The effect of group composition, social processes, and cognition. In Mumford MD (Ed.), Handbook of organizational creativity (pp. 327–357). Waltham, MA: Elsevier.

Runco MA (1991). The evaluative, valuative, and divergent thinking of children. The Journal of Creative Behavior, 25(4), 311–319. 10.1002/j.21626057.1991.tb01143.x

Runco MA (1999). Divergent thinking. In Runco MA, & Pritzker SR(Eds.). Encyclopedia of creativity (pp. 577–582). Academic Press.

Runco MA (2004). Personal creativity and culture. In Lau S, Hui ANN, & Ng GYC (Eds.), Creativity: When east meets west (pp. 9–21). World Scientific Publishing Company.

Runco MA (2020). Idea evaluation. In Runco MA, & Pritzker SR (Eds.), Encyclopedia of creativity (3rd ed., pp. 607–611). Academic Press.

Runco MA, & Chand I (1994). Problem finding, evaluative thinking, and creativity. In Runco MA(Ed.), Problem finding, problem solving, and creativity (pp. 40–76). Norwood, NJ: Ablex.

Runco MA, & Chand I (1995). Cognition and creativity. Educational Psychology Review, 7, 243–267. 10.1007/BF02213373

Runco MA, & Charles RE (1993). Judgments of originality and appropriateness as predictors of creativity. Personality and Individual Differences, 15(5), 537–546. 10.1016/0191-8869(93)90337-3

Runco MA, & Dow GT (2004). Assessing the accuracy of judgments of originality on three divergent thinking tests. The International Journal of Thinking & Problem Solving, 14(2), 5–14.

Runco MA, & Jaeger GJ (2012). The standard definition of creativity. Creativity Research Journal, 24(1), 92–96. 10.1080/10400419.2012.650092

Runco MA, & Smith WR (1992). Interpersonal and intrapersonal evaluations of creative ideas. Personality and Individual Differences, 13(3), 295–302. 10.1016/0191-8869(92)90105-X

Runco MA, & Vega L (1990). Evaluating the creativity of children's ideas. Journal of Social Behavior & Personality, 5(5), 439–452.

Rusticus SA, & Hubley AM (2006). Measurement invariance of the Multidimensional Body-Self Relations Questionnaire: Can we compare across age and gender? Sex Roles, 55, 827–842. doi:10.1007/s11199-006-9135-7

Saad G, Cleveland M, and Ho L (2015). Individualism–collectivism and the quantity versus quality dimensions of individual and group creative performance. J. Bus. Res 68, 578–586. 10.1016/j.jbusres.2014.09.004

Said-Metwaly S, Van den Noortgate W, & Barbot B (2020). Torrance test of creative thinking-verbal, Arabic version: Measurement invariance and latent mean differences across gender, year of study, and academic major. Thinking Skills and Creativity. 10.1016/j.tsc.2020.100768 (2020).

Shao Y, Zhang C, Zhou J, Gu T, & Yuan Y (2019). How does culture shape creativity? A mini-review. Frontiers in Psychology, 10, 1219. [PubMed: 31191410]

Silvia PJ (2008). Discernment and creativity: how well can people identify their most creative ideas? Psychology of Aesthetics, Creativity and the Arts, 2, 139–146. 10.1037/1931-3896.2.3.139

Simonton DK (1988). Scientific genius: A psychology of science. New York: Cambridge University Press.

Simonton DK (1999). Creativity as blind variation and selective retention: Is the creative process Darwinian9 Psychological Inquiry, 10, 309–328.

Stark S, Chernyshenko OS, & Drasgow F (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. Journal of Applied Psychology, 91, 1292–1306. https://psycnet.apa.org/doi/10.1037/0021-9010.91.6.1292 [PubMed: 17100485]

Stemler SE, & Kaufman JC (2020). Are creative people better than others at recognizing creative work?. Thinking Skills and Creativity, 38, 100727. 10.1016/j.tsc.2020.100727 [PubMed: 32952738]

Sun DZ, Li SD, Liu Y, Zhang Y, Mei R, & Yang MH (2013). Differences in the origin of philosophy between Chinese medicine and western medicine: exploration of the holistic advantages of Chinese medicine. Chinese Journal of Integrative Medicine, 19(9), 706–711. 10.1007/s11655-013-1435-5 [PubMed: 23975136]

Thompson MS, & Green SB (2006). Evaluating between-group differences in latent variable means. In Hancock GR, & Mueller GR (Eds.), Structural equation modeling: A second course (pp. 119–169). Information Age Publishing.

Vandenberg RJ, & Lance CE (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. Organizational Research Methods, 3(1), 4–70.

Van de Schoot R, Lugtig P, & Hox J (2012). A checklist for testing measurement invariance. European Journal of Developmental Psychology, 9(A), 486–492.

Wallach M, & Kogan N (1965). Modes of thinking in young children. New York: Holt, Rinehart, & Winston.

Wallas G (1926). The art of thought. New York: Harcourt Brace.

Wang X, Sun H, Zhang A, Sun W, Wang P, & Wang Z (2011). Potential role of metabolomics approaches in the area of traditional Chinese medicine: as pillars of the bridge between Chinese and Western medicine. Journal of Pharmaceutical and Biomedical Analysis, 55(5), 859–868. 10.1016/j.jpba.2011.01.042 [PubMed: 21353755]

Wicherts JM, Dolan CV, & Hessen DJ (2005). Stereotype threat and group differences in test performance: A question of measurement invariance. Journal of Personality and Social Psychology, 89, 696–716. 10.1037/0022-3514.89.5.696 [PubMed: 16351363]

Wilson J, Ward C, Fetvadjiev VH, & Bethel A (2017). Measuring cultural competencies: The development and validation of a revised measure of sociocultural adaptation. Journal of Cross-Cultural Psychology, 48(10), 1475–1506. 10.1177/0022022117732721

Wilson RC, Guilford JP, & Christensen PR (1953). The measurement of individual differences in originality. Psychological Bulletin, 50(5), 362–370. 10.1037/h0060857 [PubMed: 13100527]

Wu AD, Zhen L, & Zumbo BD (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. Practical Assessment, Research, and Evaluation, 12(1), 3. 10.7275/mhqa-cd89

Xie G, & Paik Y (2019). Cultural differences in creativity and innovation: are Asian employees truly less creative than western employees? Asia Pac. Bus. Rev 25, 123–147. 10.1080/13602381.2018.1535380

Xu L, & Barnes LL (2011). Measurement invariance of scores from the Inventory of School Motivation across Chinese and US college students. International Journal of Testing, 11(2), 178–210. 10.1080/15305058.2010.542357

Yuan KH, & Bentler PM (2006). Mean comparison: Manifest variable versus latent variable. Psychometrika, 71(1), 139–159. 10.1007/s11336-004-1181-x

Zha P, Walczyk JJ, Griffith-Ross DA, Tobacyk JJ, & Walczyk DF (2006). The impact of culture and individualism–collectivism on the creative potential and achievement of American and Chinese adults. Creativity Research Journal, 18(3), 355–x366. 10.1207/s15326934crj1803_10

Zhang ZS, Hoxha L, Aljughaiman A, Arënliu A, Gomez- Arizaga MP, Gucyeter S, … & Ziegler A (2021). Social environmental factors and personal motivational factors associated with creative achievement: A cross- cultural perspective. The Journal of Creative Behavior, 55(2), 410–432.

Highlights

- Measurement invariance of two evaluative measures was examined cross-culturally.
- A two-factor model was supported in both the American and Chinese groups.
- The model satisfied configural, weak and partial strong invariance cross-culturally.
- Latent mean comparisons suggested higher evaluative skill of Americans on Uses task.

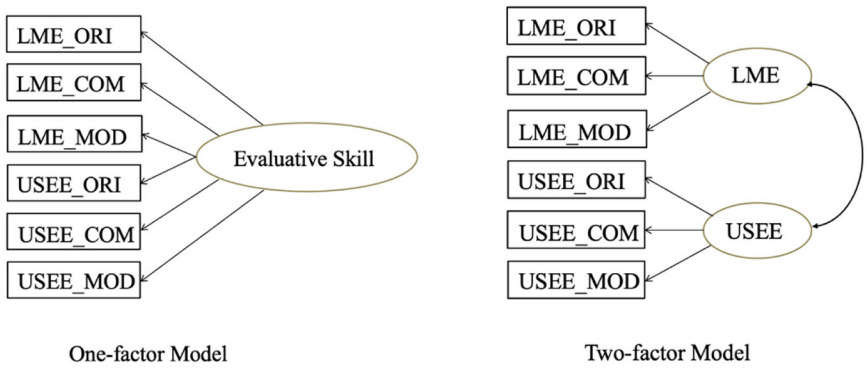

Figure 1. Theoretical Models to be Tested

Note. LME_ORI, LME_COM, LME_MOD refer to the average scores obtained over the four highly original ideas, four common ideas, and four moderately original ideas evaluated on Line Meanings; USEE_ORI, USEE_COM, USEE_MOD refer to the average scores obtained over the four highly original ideas, four common ideas, and four moderately original ideas evaluated on Uses. LME = Line Meanings Evaluative Accuracy; USEE = Uses Evaluative Accuracy.

Table 1

Means, Standard Deviations, and Correlations between Evaluative Accuracy Scores

	American		Chinese							
	M	***SD***	***M***	***SD***	**1**	**2**	**3**	**4**	**5**	**6**
LME_ORI	.43	.16	.50	.17	1	.457**	.549**	.423**	.464**	.509**
LME_COM	.47	.14	.44	.14	.501**	1	.543**	.329**	.438**	.469**
LME_MOD	.47	.16	.42	.15	.600**	.518**	1	.404**	.445**	.476**
USEE_ORI	.41	.16	.44	.16	.439**	.446**	.542**	1	.595**	.604**
USEE_COM	.31	.16	.36	.16	.292**	.377**	.374**	.603**	1	.639**
USEE_MOD	.39	.16	.47	.17	.334**	.419**	.466**	.654**	.669**	1

Note. American sample $n = 341$; Chinese sample $n = 345$; correlations in the American sample were below the diagonal; correlations in the Chinese sample were above the diagonal; LME_ORI, LME_COM, LME_MOD refer to the average scores obtained over the four highly original ideas, four common ideas, and four moderately original ideas evaluated on Line Meanings; USEE_ORI, USEE_COM, USEE_MOD refer to the average scores obtained over the four highly original ideas, four common ideas, and four moderately original ideas evaluated on Uses.

** $p < .01$.

Table 2

Model Comparison of the One- and Two-Factor Models

Model	No. of factors	χ^2	*df*	CFI	TLI	RMSEA	SRMR	$\Delta\chi^2$
Configural	One	198.873	18	.891	.819	.171	.063	
Configural	Two	43.441	16	.984	.969	.071	.029	155.432

Note. *df* = degree of freedom; CFI = Comparative fit index; TLI = Tucker Lewis index; RMSEA = Root mean square error of approximation; SRMR = Standardized root mean square residual.

Table 3

Configural Invariance Model Parameter Estimates across Two Groups

	LME by			USEE by		
	LME_ORI	**LME_COM**	**LME_MOD**	**USEE_ORI**	**USEE_COM**	**USEE_MOD**
American	.723***	.680***	.757***	.730***	.789***	.826***
Chinese	.719***	.674***	.812***	.811***	.763***	.829***

Note. Coefficients are standardized. LME_ORI, LME_COM, LME_MOD reter to the average scores obtained over the four highly original ideas, four common ideas, and four moderately original ideas evaluated on Line Meanings; USEE_ORI, USEE_COM, USEE_MOD refer to the average scores obtained over the four highly original ideas, four common ideas, and four moderately original ideas evaluated on Uses.

*** $p < .001$.

Table 4

Model Comparisons across Two Groups

Model	χ^2	*df*	CFI	RMSEA	SRMR	Compare with	$\Delta\chi^2$	Δdf	ΔCFI	ΔRMSEA
Configural	43.441	16	.984	.071	.029					
Weak	46.757	20	.984	.062	.035	configural	3.316	4	.000	−.009
Strong	175.927	24	.909	.136	.072	weak	129.170***	4	−.075	.074
Partial strong	62.975	22	.975	.074	.041	weak	16.218***	2	−.009	.012

Note. *df* = degree of freedom; CFI = Comparative fit index; RMSEA = Root mean square error of approximation; SRMR = Standardized root mean square residual. Partial strong: The model with intercepts of LME_ORI and LME_MOD freely estimated between groups (i.e., testing the invariance of the Uses evaluation task).

*** $p < .001$.