Creativity Research Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hcrj20

A Reliability Generalization Meta-Analysis of the Creative Achievement Questionnaire

Süreyya Yörük & Sedat Sen

To cite this article: Süreyya Yörük & Sedat Sen (2022): A Reliability Generalization Meta-Analysis of the Creative Achievement Questionnaire, Creativity Research Journal, DOI: 10.1080/10400419.2022.2148073

To link to this article: https://doi.org/10.1080/10400419.2022.2148073

Published online: 05 Dec 2022.

Submit your article to this journal

Article views: 120

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=hcrj20

CREATIVITY RESEARCH JOURNAL
https://doi.org/10.1080/10400419.2022.2148073

Routledge
Taylor & Francis Group

Check for updates

A Reliability Generalization Meta-Analysis of the Creative Achievement Questionnaire

Süreyya Yörük [a] and Sedat Sen [b]

[a]Marmara University; [b]Harran University

ABSTRACT
The Creative Achievement Questionnaire (CAQ) is widely used to measure the creative achievement levels of individuals. Previous studies reported a varying range of reliability coefficients for the CAQ. To this date, no study has investigated the variability in the reliability coefficients of the CAQ. A random-effects reliability generalization meta-analysis, heterogeneity analyses, and moderator analyses were conducted in the present study to examine the variability in the reliability coefficients. The mean alpha value for the total CAQ score was estimated to be .765 (95% CI: .708-.811). The *Q* test ($Q_{(33)}$ = 1716.648, p < .0001) and I^2 (98.74%) values denoted statistically significant heterogeneity among the alpha coefficients. These findings indicate that reliability is a property of CAQ scores and that the reliability of scores changes significantly across samples. The moderator analyses showed that the reliability coefficients were not impacted by the moderator variables considered in this study. However, certain moderator variables that were not included in the analyses may have impacted the reliability coefficients. The findings of the present study imply that the CAQ can be used for exploratory research.

ARTICLE HISTORY
Received February 10, 2022

Since the first day of COVID-19, people all over the world have faced different types of new problems that needed to be solved creatively. Creativity not only makes it possible for us to overcome unprecedented problems, such as the ones originating from COVID-19, but also allows us to deal with less complex challenges in life, such as planning a weekend with friends. People display various levels of creative achievement to live a better life.

Carson, Peterson, and Higgins (2005) defined creative achievement as "the sum of creative products generated by an individual in the course of his or her lifetime" (p. 37). Creative products refer to one's work that is both original and useful (Barron, 1955; Mumford, 2003; O'quin & Besemer, 1999; Runco & Jaeger, 2012; Stein, 1953). This means that – whether it is a product of an artist, a layperson, a professional, or a scientist – a creative product must be unique or novel and must offer some type of functionality or value.

Creative products come in different levels. According to Kaufman and Beghetto (2009), a product may be an example of mini-c, little-c, Pro-c, or Big-C creativity. Different levels of creative products bear different levels of creative achievements. Hence, researchers exploit one's creative products to assess his or her creative achievements and to determine his or her level of creativity (Silvia, Wigert, Reiter-Palmon, & Kaufman, 2012). Assessment of creative achievement allows researchers to examine creativity in different aspects, such as the domain specificity of creativity (Carson et al., 2005; Silvia et al., 2012) or the relationship between creativity and other variables (e.g., intelligence and culture; Badibanga, 2013; Kaufman, 2009; Polner, Nagy, Takáts, & Kéri, 2015).

Different instruments were developed for the purpose of measuring creative achievement, including the Creative Behavior Inventory (Hocevar, 1979), the Creative Achievement Scale (Ludwig, 1992), the Lifetime Creativity Scales (Richards, Kinney, Benet, & Merzel, 1988), and the Creative Achievement Questionnaire (CAQ, Carson et al., 2005). The results on Google Scholar show that the CAQ is the most commonly used instrument. Previous research examined the CAQ from different aspects, such as internal structure, predictive validity, internal consistency, and temporal stability (e.g., Carson et al., 2005; Hezel & Hooley, 2014; Ruiter & Johnson, 2015). However, the variability in the reliability of CAQ scores has not been investigated. In an attempt to fill this gap in the literature, the present study focuses on the variation in the reliability coefficients that were estimated in different applications of the CAQ.

CONTACT Süreyya Yörük sureyya.yoruk@marmara.edu.tr Faculty of Education, Marmara University, İstanbul 34722, Turkey
© 2022 Taylor & Francis Group, LLC

The Creative Achievement Questionnaire

The CAQ is a self-reported checklist that measures the level of creative achievement in 10 domains (Carson et al., 2005). Those domains are as follows: visual arts, music, creative writing, dance, drama, architecture, humor, scientific discovery, invention, and culinary. The CAQ is composed of three parts and 96 items in total.

In Part 1, individuals put a checkmark next to their perceived talent area(s) (Carson et al., 2005). Thirteen talent areas are offered, such as music, team sports, and invention. To score the items, yes-no scoring is used – note that yes-no scoring is converted to 1–0 scoring for analysis purposes (i.e., a checkmarked item receives 1 point).

Part 2 consists of 80 items in total. Individuals respond to eight items for each domain and put a checkmark next to those items that match their past achievements (Carson et al., 2005). The first item in each domain indicates that an individual has no training, experience, or recognized creative talent in that domain (e.g., I have no training or recognized talent in this area; Carson et al., 2005). The eighth item states that an individual's creative achievements were recognized (e.g., published, cited by others, reviewed, or sold) or criticized nationally (Carson et al., 2005). If an item is checkmarked, the item number is awarded as the item score. Some items that are identified with an asterisk require individuals to specify the number of times a particular achievement was displayed (Carson et al., 2005). To compute the score for those items, the item number is multiplied by the number of times the achievement was displayed. The total CAQ score is the sum of all the points given to the items in Part 2 and ranges from 0 to an undetermined number with no upper limit.

In Part 3, individuals respond to three items. Each item is a statement that focuses on a certain characteristic of creative people (i.e., being perceived as more creative than others and being seen as "having an artistic temperament" or "an absent-minded professor"; Carson et al., 2005, p. 50). These items are scored in a binary fashion during the analysis (i.e., a checkmarked item receives 1 point).

The items for each domain are ordered hierarchically, and each item corresponds to a certain level of creative achievement (Carson et al., 2005). There is no validated short version of the CAQ. However, a few studies focused on certain domains due to the scope of the research and did not use all the items (e.g., Bojner-Horwitz, Lennartsson, Theorell, & Ullén, 2015; Kuckelkorn, de Manzano, & Ullén, 2021; Lu, 2017; Mosing et al., 2015). Additionally, the first and third parts of the CAQ were not given to the participants in some studies (e.g., Zhu, Shang, Jiang, Pei, & Su, 2019).

The CAQ was translated into several languages, such as German (Diedrich et al., 2018), Italian (Salvi et al., 2020), and Chinese (Deng, Wang, & Zhao, 2016). However, only the Chinese version of the instrument was standardized in a reliability and validity study (see Zhanqi & Xingli, 2020). The rest of the adaptations were mere translations (e.g., de Manzano & Ullén, 2018; Form, Aue, & Kaernbach, 2019; Kéri, 2011; Rahmani, Moghaddam, & Aarabi, 2020; Vellante et al., 2011; Wang, Ho, Cheng, & Cheng, 2014).

The purpose of the CAQ is to distinguish "highly creative from less creative individuals across a variety of samples and a variety of domains" (Carson, 2001, p. 5). Carson et al. (2005) did not specify for which populations the CAQ was developed. In the original study (Carson, 2001), the reliability and validity of scores were examined with undergraduate and graduate students. The majority of studies were also conducted with undergraduate students and adults (e.g., Silvia et al., 2012). The CAQ was administered to high school students in a few studies (e.g., Kaufman, 2009). In one study, individuals as young as 13 years of age were the participants (Du et al., 2019). As seen, the target populations of the CAQ include individuals from the general population.

The reliability and validity of CAQ scores were examined in different studies. Carson et al. (2005) addressed validity. Scores on the CAQ and scores on colleges were correlated to examine predictive validity. The zero-order coefficient value was .59. However, the coefficient value was estimated to be .65 after aesthetic appeal was controlled. With regard to convergent validity, CAQ scores were moderately correlated with openness to experience (r = .42), flexibility (r = .34), originality (r = .32), and total divergent thinking score (r = .29). With respect to discriminant validity, scores on the CAQ and scores on the Vocabulary and Block Design subtest of the Wechsler Adult Intelligence Test-Revised (Wechsler, 1981) had a weak correlation (r = .14). Finally, Carson et al. (2005) identified a three-factor structure that explained 48.8% of the variance. However, the authors tended to make use of the two-factor structure that explained 33.5% of the variance because "it [is] useful to employ the arts versus science dichotomy" (Carson, 2001). The arts factor includes drama, writing, humor, music, visual arts, and dance, whereas the science factor includes invention, science, and culinary domains. Nevertheless, several studies treated the CAQ as a single-factor instrument (e.g., Hezel & Hooley, 2014; Ruiter & Johnson, 2015).

With respect to reliability, Carson et al. (2005) reported a coefficient value of .81 for test-retest reliability. In the same study, Cronbach's alpha value was calculated to be .96 for the entire instrument. The alpha values for the 10 domains ranged from .70 to .87. Finally, the split-half reliability coefficient of the first 48 items was .92 and for the second 48 items was .91. Similar reliability coefficients were reported in several other studies (see Conover, 2015; Ruiter & Johnson, 2015; von Stumm & Scott, 2019). The lowest value was .19 (Carruthers, 2016) and the highest value was .96 (Ruiter & Johnson, 2015) for the total score.

Purpose of the present study

Previous studies reported a varying range of reliability coefficients for the CAQ. To the authors' knowledge, no study has focused on explaining the variability in the reliability coefficients of the CAQ. Evidently, why the reliability coefficients vary across samples needs to be examined. This particular examination can be achieved through the reliability generalization (RG: Vacha-Haase, 1998) approach (see Sen, 2022 for an example in creativity). RG analyses can indicate whether scores on an instrument are reliable across samples and study characteristics (Sánchez-Meca et al., 2021). Additionally, the RG methodology can be used to estimate the overall reliability coefficient of an instrument and to examine the variability in the reliability coefficients (Vacha-Haase, 1998).

One of the purposes of the present study was to conduct an RG meta-analysis to investigate the variability in the reliability of CAQ scores. To achieve this goal, the pooled score reliability coefficient was estimated. Afterward, the relationship between study characteristics and the overall reliability coefficient was examined. Because most of the studies reported Cronbach's alpha, the meta-analysis was conducted with the alpha values. It should be noted that the alpha values were the ones calculated across the 10 CAQ domains for the total score. Additionally, certain characteristics of the samples (e.g., sample size, mean age) and the instrument (e.g., application in different languages) were analyzed to explore whether those characteristics could explain the variability in the reliability coefficients.

Other purposes were to estimate the reliability induction rate for the CAQ and to compare the characteristics of the inducing studies with the reporting ones. The majority of the studies either reported the reliability coefficients estimated in previous studies or mentioned no reliability coefficient for the CAQ. The former is known as reliability induction by report, whereas the latter is known as reliability induction by omission (Sánchez-Meca et al., 2021). Both types of induction are questionable because reliability is treated as a fixed property of test scores when no reliability coefficient is estimated for a sample. In order to examine reliability induction, previous studies on the CAQ were categorized based on reporting, mentioning, and omitting Cronbach's alpha.

The research questions of the present study are as follows:

(1) What is the overall reliability coefficient of CAQ scores?
(2) How does the overall reliability coefficient change as a function of study characteristics?
(3) What are the rates for reliability induction by report and by omission of the CAQ?
(4) What are the characteristics of the inducing studies compared to those that reported Cronbach's alpha?

Method

Selection criteria

The authors conducted an internet search to find all available empirical studies that used the CAQ. Because there is no validated short version of the CAQ, studies that used the original version were considered. Some studies did not use all 96 items, but the alpha coefficients reported by those studies were also included in the analyses (see Table 2 for the number of items in each study). Some studies were not written in English, and they were excluded. In essence, studies that were empirical and written in English were selected for inclusion. Studies that had been conducted since 2005 were obtained. This specific year was chosen because 2005 was the year when Carson et al. (2005) published the article on the CAQ.

We aimed to conduct a reliability generalization meta-analysis for temporal stability and internal consistency. Thus, the test-retest, split-half, and Cronbach's alpha reliability coefficients were considered. However, the test-retest and split-half reliability coefficients were reported by only one study. Thus, only the alpha values were used to conduct the analyses. There was no restriction with respect to the characteristics (e.g., age, gender, ethnicity, and location) of the samples for inclusion.

Search strategies

The internet search was conducted in the following databases: PsycARTICLES, Proquest, PsycINFO, ERIC, Web of Science, and Academic Search Complete. Additionally, Google Scholar was exploited to access additional studies that were not available in these databases. The strings used were "Creative Achievement Questionnaire," "CAQ," and "CAQ and Carson, Peterson, and Higgins." "Creativity Achievement Questionnaire" was also used because several researchers mistakenly used this string in their papers. These strings were searched in the entire paper, including the title, abstract, and main text.

In addition to the previously mentioned databases and Google Scholar, four highly reputable journals were scanned for additional studies. Those journals were *Creativity Research Journal, Journal of Creative Behavior, Thinking Skills and Creativity*, and *Psychology of Aesthetics, Creativity, and the Arts*. As of January 1st 2022, the internet search listed 2731 papers that mentioned the CAQ – note that this number includes books, duplicates of the same papers, and non-English studies as well (see Figure 1).

Data extraction

After the books, duplicates, and non-English studies were removed, 759 studies (articles, theses, or dissertations) remained. We skimmed all 759 studies and detected that the CAQ was used in 257 studies. The rest of the studies did not use the CAQ and just cited the original study (Carson et al., 2005). Therefore, the remaining 502 studies were excluded.

Not all 257 studies reported reliability coefficients. Only 37 studies estimated and reported reliability coefficients. Two studies (Ahmetoglu, 2014; Silvia, Kaufman, Reiter-Palmon, & Wigert, 2011) were conducted with the same sample as another study that used the CAQ. Three studies focused on one domain and used a very low number of items (Chai & Fan, 2016; Du et al., 2019; Onwe, 2019). Thus, those studies were excluded as well. Several studies either reported the reliability coefficients estimated in previous studies or mentioned nothing about the reliability of CAQ scores. These were coded as induction by report and induction by omission, respectively. It should be noted that the inducing studies were also excluded, as suggested by Sánchez-Meca et al. (2021).

We sent an e-mail to the corresponding authors of the inducing studies. The majority of the corresponding

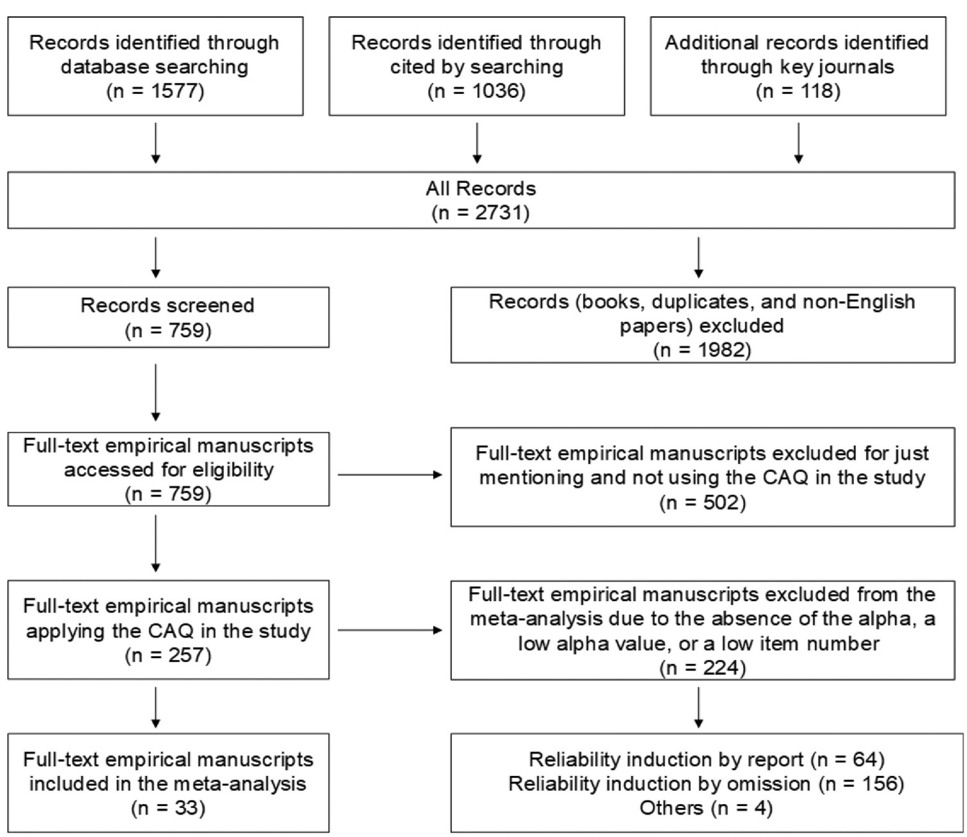

Figure 1. REGEMA flow diagram.

Table 1. Coding system for studies.

Label	Type	Coding method
Study ID	Cat.	Arbitrary ID assigned to each sample within a study
Authors	Cat.	Authors' surnames noted
Year	Cont.	Publication year noted
Title	Cat.	Title of the study noted
Publication type	Cat.	0 = for non-published (i.e., thesis); 1 = published (i.e., article)
Number of items	Cont.	Number of CAQ items used as a sole scale
Test version	Cat.	0 = Long (96 items); 1 = Reduced (<96 items)
Sample size	Cont.	Valid sample size for the internal consistency estimate
Alpha estimate	Cont.	Cronbach's alpha coefficient
Mean age	Cont.	Average age (in years) reported for the sample
Age SD	Cont.	Standard deviation of the age (in years) reported for the sample
Gender	Cont.	Proportion of females to total sample size
Continent	Cat.	0 = North America; 1 = Europe; 2 = Asia
Country	Cat.	0 = sample collected in another country; 1 = sample collected in the U.S.
Language	Cat.	0 = non-English; 1 = English.
Population type	Cat.	0 = general; 1 = students
Education level	Cat.	0 = graduate; 1 = undergraduate; 2 = high school; 3 = mixed
Culture	Cat.	0 = Western; 1 = Eastern
Ethnicity	Cat.	0 = Caucasian; 1 = Asian; 2 = Mixed

Note. CAQ = Creative Achievement Questionnaire; Cat = categorical; Cont = continuous

authors did not e-mail back or responded that they could not calculate and share the reliability coefficients. Five of the corresponding authors shared their estimations. This process yielded an additional five reliability coefficients. Overall, 37 studies whose reliability coefficients were obtained were considered for the analyses.

Reliability estimate

In this RG study, Cronbach's alpha was considered for the analyses as it was reported in the majority of the samples (89.48%). The rest of the studies (Balta, Subotić, & Pavićević, 2019; Carson et al., 2005; Lebuda, Zielińska, & Karwowski, 2021) reported different reliability indices, including the test-retest (2.63%), Omega (2.63%) and H (5.26%) coefficients. A separate meta-analysis could not be conducted for each type of reliability coefficient due to a low number of coefficients. Therefore, only 35 alpha coefficients obtained from 34 studies were retained for the RG analyses. One study (Carruthers, 2016) reported a small alpha value (.186) that can be considered an outlier. Thus, the final sample included 34 alpha values after the outlier was excluded.

Earlier studies have shown that most of the reliability indices have a highly skewed distribution. Thus, we transformed the alpha values to Bonett's (2002) T-values. We used the formula $L_i = \ln(1-|\alpha_i|)$ to normalize the distribution and stabilize the sampling variance.

Statistical analyses

The possibility of publication bias and other types of biases were assessed before proceeding to the main analyses. Egger's regression test (Egger, Smith, Schneider, & Minder, 1997) as well as Begg and Mazumdar's rank correlation test (Begg & Mazumdar, 1994) were used to investigate publication bias – statistically non-significant *p*-values indicate the lack of publication bias. In addition, fail-safe *N* numbers proposed by Rosenthal (1979) and Rosenberg (2005) were calculated – high *N*s indicate the lack of publication bias. A funnel plot was also created to detect whether the alpha coefficients had a symmetrical shape – the symmetrical funnel plot indicates the lack of publication bias. We also identified reliability induction by omission and induction by report, as recommended in the REGEMA checklist (Sánchez-Meca et al., 2021).

To obtain the pooled mean of the alpha coefficient, a random-effects RG meta-analysis (Vacha-Haase, 1998) was conducted with inverse variance weights. The heterogeneity among the alpha coefficients was checked with *Q*-statistics and I^2 values. Statistically significant *Q*-statistics and large I^2 values (>75%; Higgins, Thompson, Deeks, & Altman, 2003) indicate heterogeneity.

The characteristics of the sample, instrument, or research affect the reliability coefficients in a study. Thus, certain moderator variables were considered in the present study (see Table 1 for the coding system).

Table 2. Summary of the 33 included studies with 34 reliability coefficients.

ID	Author(s)	*k*	*N*	α	Country	Popl.	Lang.	Education level	Female %	Mean Age	Age SD	Ethnicity
1	Carson et al., 2005	96	117	.960	US	Students	English	Undergad	43.50	20.10	1.60	Mixed
2	Kaufman, 2009*	96	177	.740	England	Students	English	Undergrad	67.70			Mixed
3	Shelton, 2010*	96	155	.540	US	General	English		61.30	42.15	12.96	Mixed
4	Keri, 2011	80	111	.830	Hungary	General	Hungarian		41.44	41.60	8.70	Caucasian
5	Ofili, 2011*	96	122	.710	US	Students	English	mixed	54.92	23.46	4.64	Mixed
6	Silvia et al., 2012	96	848	.600	US	Students	English	Undergrad	76.00	22.90	6.63	Mixed
7	Silvia & Nusbaum, 2012	96	188	.820	US	Students	English	Undergrad	73.94			Mixed
8	Lucas, van der Wijst, Curşeu, & Looman, 2013	40	1543	.625	Netherlands	Students	Dutch	mixed	74.60	17.79	1.50	Caucasian
9	Hezel & Hooley, 2014	96	80	.470	US	General	English		57.50	26.40	11.30	
10	Conover, 2015*	96	833	.810	US	General	English		52.70	34.30	13.01	Mixed
11	Dostal et al., 2015	70	1112	.650	Czech Rep.	Students	Czach	mixed	75.18	22.77	2.12	Caucasian
12	Ruiter & Johnson, 2015	96	297	.960	US	Students	English	Undergrad	67.20			Mixed
13	Ahmetoglu et al., 2015	72	210	.760	UK	General	English		55.71	2.22	0.80	Caucasian
14	Deng et al., 2016_1	32	193	.840	US	Students	English	Undergrad	59.90	21.46	1.37	Caucasian
15	Deng et al., 2016_2	32	185	.870	China	Students	Chinese	Undergrad	68.50	22.67	2.42	Asian
16	Diedrich et al., 2018	96	541	.647	Austria	General	GermanEnglish		67.20	26.77		Mixed
17	Forgeard & Benson, 2017	96	512	.548	US	Students	English	Undergrad	64.50	18.96	0.78	Mixed
18	Grohman, Ivcevic, Silvia, & Kaufman, 2017	96	131	.580	US	Students	English	Undergrad	64.88	19.37	1.85	Mixed
19	Willard, 2017*	72	209	.736	US	Students	English	Undergrad	76.55	19.48		Mixed
20	Falavarjani, 2018	96	328	.640	Malaysia	Students	Persian	Grad	42.00	32.00	5.73	Middle East
21	Han et al., 2018	96	286	.719	China	Students	Chinese	Undergrad	50.34	20.00		Asian
22	Salvi et al., 2020	96	153	.460	Italy	Students	Italian	Undergrad	62.09	24.80	4.90	Caucasian
23	Specker et al., 2018	93	668	.600	Austria	Students	German	Undergrad	68.70	23.85	9.20	Caucasian
24	Mendes et al., 2019	96	79	.670	Germany	General	German		48.45	34.00	16.00	Caucasian
25	von Stumm & Scott, 2019	96	128	.920	UK	General	English		57.81	29.51	8.60	Mixed
26	Sutu, Serrano, Schultz, Jackson, & Damian, 2019	70	1344	.810	US	Students	English	Undergrad	68.00	19.52	2.04	Mixed
27	Zhu et al., 2019	64	588	.880	China	Students	Chinese	High school	69.04	16.56	0.70	Asian
28	Duan et al., 2020	96	70	.716	China	Students	Chinese	Undergrad	0.00	19.24	1.38	Asian
29	Loh & Lim, 2020	96	104	.683	Singapore	Students		Undergrad	61.54	21.60	1.70	Asian
30	Rodet, 2021	96	111	.710	US	General	English		57.00			
31	Jacops et al., 2021	96	559	.700	Netherlands	Students	Dutch and English	Undergrad				Caucasian
32	Lebuda et al., 2021	96	1669	.760	Poland	General	Polish		63.00	56.00	0.00	Caucasian
33	Patil, Madathil, & Huang, 2021	96	53	.910	Taiwan	General	Chinese		45.00	43.00		Asian
34	Verger et al., 2021	47	168	.780	France	General	French		81.55	35.10	13.82	Caucasian

Note. *k* = number of items; *N* = sample size; α = Cronbach's alpha; *thesis. Popl. is population. Lang. is language.

Because different numbers of items were administered in previous studies, item number was one of the moderator variables. The R package "metafor" (Viechtbauer, 2010) was used for the analyses, including the reliability generalization meta-analysis (Vacha-Haase, 1998) as well as the heterogeneity, moderator, and publication bias analyses.

Results

Description of the characteristics of the studies

Thirty-four coefficient alpha values obtained from 33 studies were included in the RG analyses. Characteristics of those studies are presented in Tables 2 and 3. As shown in Table 2, the years of studies ranged from 2005 to 2021. The selected studies consisted of journal articles (85.30%) or theses (14.70%). The studies were conducted in several countries, including the U.S. (41.20%), China (11.80%), the UK (8.80%), Netherlands (5.90%), Austria (5.90%), Czech Republic (2.90%), France (2.90%), Germany (2.90%), Hungary (2.90%), Italy (2.90%), Malaysia (2.90%), Poland (2.90%), Singapore (2.90%), and Taiwan (2.90%). The studies were conducted in three continents: Asia (20.60%), Europe (38.20%) and North America (41.20%). The studies can be categorized as East (20.60%) and West (79.40%) in terms of culture. The ethnicities of the samples were Asian (18.75%), Caucasian (34.38%), Middle Eastern (3.12%), and mixed (43.75%).

In total, there were 13,872 participants (M = 408, SD = 437.58). The mean age of the participants ranged from 16.56 to 56.00 (M = 26.97, SD = 9.55). The standard deviation of the participants' age was between 0 and 16 (M = 5.54, SD = 4.95). The percentages of females ranged from 0 to 81.55 (M = 59.93, SD = 15.06). The participants were either students (64.70%) or individuals from the general population (35.30%). The educational levels of the student samples were undergraduate (77.27%),

Table 3. A summary of study characteristics of the included studies.

	k	%
Publication type (*n* = 34)		
Article	29	85.30
Thesis	5	14.70
Test length (*n* = 34)		
96-item	23	67.60
<96-item	11	32.40
Country (*n* = 34)		
US	14	58.80
Non-US	20	41.20
Continent (*n* = 34)		
Asia	7	20.60
Europe	13	38.20
North America	14	41.20
Culture (*n* = 34)		
East	7	20.60
West	27	79.40
Language (*n* = 33)		
English	17	51.52
Non-English	14	42.42
Mixed	2	6.06
Population type (*n* = 34)		
General	12	35.30
Students	22	64.70
Education level (*n* = 22)		
Undergraduate	17	77.27
Graduate	1	4.55
High school	1	4.55
Mixed	3	13.63
Ethnicity (*n* = 32)		
Asian	6	18.75
Caucasian	11	34.38
Middle East	1	3.12
Mixed	14	43.75
	M	*SD*
% of females (*n* = 33)	59.93	15.06
# of items (*n* = 34)	84.71	19.99
Sample size (*n* = 34)	408.00	437.59
Mean age (*n* = 28)	26.97	9.55
Age SD (*n* = 24)	5.35	4.93

graduate (4.55%), high school (4.55%), and mixed (i.e., high school and university students; 13.63%).

The English version of the CAQ was administered in 51.52% of the studies. Translated versions, including Chinese, Czech, Dutch, French, German, Hungarian, Italian, Persian, and Polish versions, were used in the rest of the studies (48.48%). The number of items used in 34 CAQ applications ranged from 32 to 96 (*M* = 84.71, *SD* = 19.99). Sixty-eight percent of the studies used the full version, whereas 32% of the studies excluded certain items. The CAQ was treated as a multi-factor instrument in six studies, and therefore, a separate alpha coefficient was estimated for each factor – note that only the highest alpha coefficient was included in the analyses.

Estimate of pooled reliability

The mean reliability value of the raw alpha coefficients was .725 (*SD* = 0.131, median = .717). The pooled

.4 | 67

.5 | 458

.6 | 00345578

.7 | 0112244668

.8 | 1123478

.9 | 1266

Figure 2. Stem and leaf plot of raw alpha values.

estimate indicates sufficient internal consistency for the CAQ. A stem and leaf plot for the raw alpha coefficients is presented in Figure 2. As shown in Figure 2, the lowest alpha value was .46 and the highest value was .96. Twenty-one samples reported reliability coefficients higher than .70. This finding indicates that CAQ scores were adequately reliable in 62% of the samples (O'Rourke, Hatcher, & Stepanski, 2005).

The random-effects model analyses were conducted with the transformed alpha values (i.e., Bonett, 2002 transformation). The pooled transformed coefficient alpha value was estimated to be .765 (95% CI: .708-.811 and PI: .167, .934) and found to be statistically significant ($p < .001$). The average transformed reliability coefficient was slightly higher than the mean of the raw reliability coefficients. The result of the *Q* test ($Q_{(33)} = 1716.648$, $p < .0001$) showed statistically significant heterogeneity among the alpha coefficients. The I^2 value was 98.74% and indicated a high amount of heterogeneity in the reliability estimates. The variability in the alpha coefficients can also be seen in the forest plot (see Figure 3).

Relationship between the moderating variables and reliability estimate

Differences among the subgroups were examined with the analog to the ANOVA approach. The subcategories of the nine categorical variables are presented in Table 4. As seen in Table 4, none of the categorical variables (publication type, continent, country, culture, ethnicity, language, population type, education level, and test version) explained a statistically significant amount of variability in the reliability estimates ($p > .05$). The statistically non-significant *p*-values showed the lack of heterogeneity in the reliability estimates among the subgroups. It should be noted that statistically non-significant results may occur due to unevenly distributed cells.

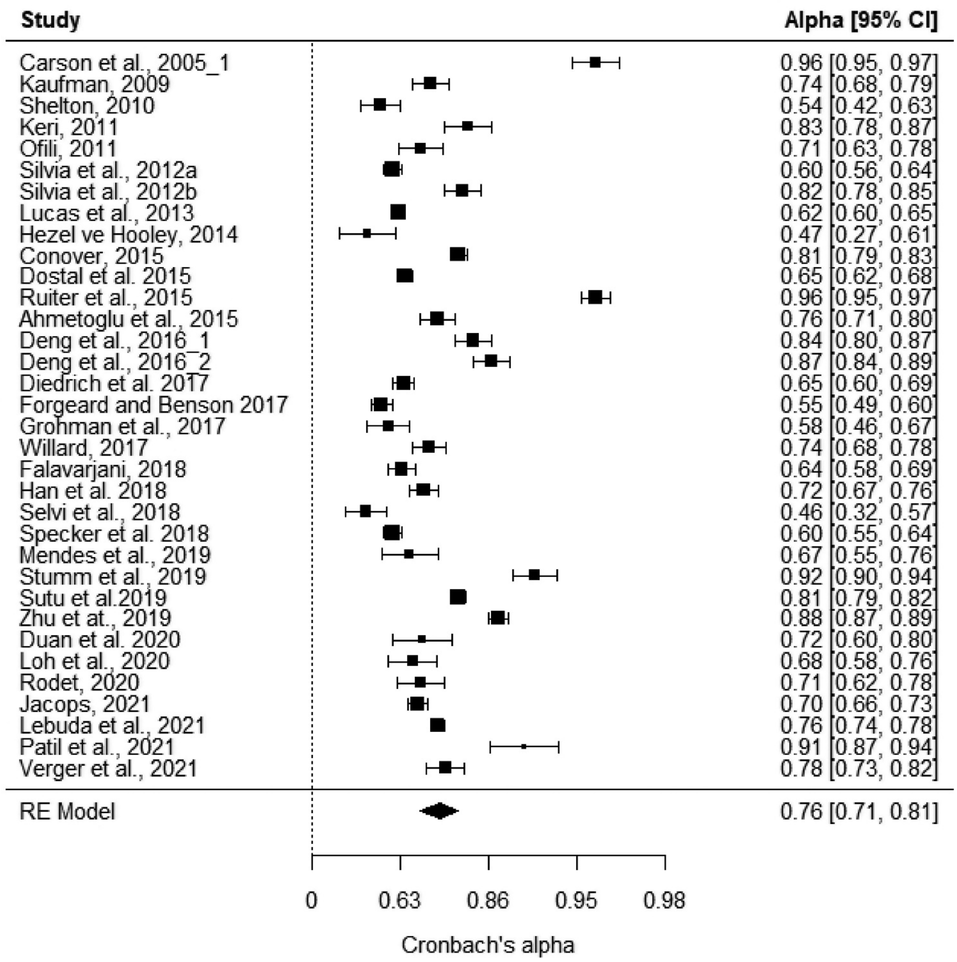

Figure 3. Forest plot for Cronbach's alpha values.

Table 4. The results of categorical moderators (Analog to the ANOVA).

Variable	Category	*k*		95%CI	Q_B	*df*	*p*
Publication Type	Published	29	.7720	[.7110, .8201]	.425	1	.515
	Non-published	5	.7200	[.5048, .8417]			
Continent	Asia	7	.7992	[.6734, .8766]	1.200	2	.549
	Europe	13	.7273	[.6115, .8087]			
	S. America	14	.7789	[.6887, .8429]			
Country	U.S.	14	.7789	[.6887, .8429]	0.207	1	.649
	non-U.S.	20	.7548	[.6735, .8158]			
Culture	East	7	.7992	[.6741, .8763]	0.509	1	.475
	West	27	.7553	[.6875, .8085]			
Ethnicity	Asian	6	.8184	[.6934, .8924]	2.408	2	.299
	Caucasian	11	.7157	[.5831, .8061]			
	Mixed	14	.7963	[.7140, .8549]			
Language	English	17	.7885	[.7086, .8465]	0.4608	1	.497
	non-English	14	.7505	[.6447, .8248]			
Population type	General	12	.7674	[.6895, .8202]	0.004	1	.947
	Students	22	.7637	[.6624, .8397]			
Education level	Undergraduate	17	.7748	[.6847, .8391]	0.338	1	.561
	Other	5	.7226	[.4856, .8504]			
Test version	96-item	23	.7577	[.6833, .8146]	0.152	1	.696
	Reduced	11	.7793	[.6758, .8498]			

Note. *k* = number of alpha coefficients, *CI* = confidence interval, *df* = degrees of freedom.

Table 5. The results of continuous moderators (meta-regression).

Moderator Variable	*k*	b_j	SE	*p*	R^2	Q_E
Year	34	−0.005	0.0288	.3878	.001	1698.122**
Mean of the age (in years)	28	0.0008	0.0126	.7233	.001	1157.457**
SD of the age (in years)	24	−0.0219	0.0269	.4163	.001	1120.807**
Female %	33	−0.0011	0.0078	.5358	.001	1622.919**

Note. k = number of alpha coefficients; b_j = unstandardized regression coefficient; SE = standard error; R^2 = proportion of variance explained; Q_E = statistic to test for residual heterogeneity, **$p < .001$

The effects of the continuous variables (year, female proportion, mean age, and *SD* of mean age) on the alpha coefficients were examined with the meta-regression approach. The results are presented in Table 5. The meta-regression analyses showed that none of the variables were statistically significant predictors of the transformed alpha coefficients ($p > .05$).

Publication bias

The publication bias was assessed with a number of methods, including the fail-safe *N* method, Egger's regression test, and Begg and Mazumdar's rank correlation test. The traditional fail-safe *N* (Rosenthal, 1979) and Rosenberg's (2005) fail-safe *N* values were found to be 131,832 and 109,635, respectively. These values were larger than $5k+10$ (i.e., 180) and indicated the lack of publication bias. Kendall's τ between the coefficient alpha values and the corresponding standard errors was positive (.130) and statistically non-significant ($p = .288$) based on Begg and Mazumdar's rank correlation test. In addition, the intercept of Egger's regression was found to be 1.273 with a statistically non-significant *p*-value (.327) based on a regression test for funnel plot asymmetry. The statistically non-significant *p*-values from these two tests indicate the lack of publication bias. The funnel plot for the alpha coefficients and standard errors is illustrated in Figure 4. As shown in Figure 4, the funnel plot has a slightly asymmetrical shape. These results indicate that publication bias was not a serious threat.

Reliability induction

To estimate the reliability induction rate, the studies were categorized based on estimating, reporting, and omitting reliability coefficients. Reliability coefficients were estimated in only 42 studies. It should be noted that nine studies were excluded from the analyses. Sixty-four studies relied on the reliability estimations of other studies – primarily the study conducted by Carson et al. (2005). One hundred and fifty-six studies mentioned nothing about reliability at all. The total reliability induction rate was 85.60%. The rate of reliability induction by report was 29.10%, whereas the rate of reliability induction by omission was 70.90%.

In the present study, only the studies that reported Cronbach's alpha for CAQ scores were included in the

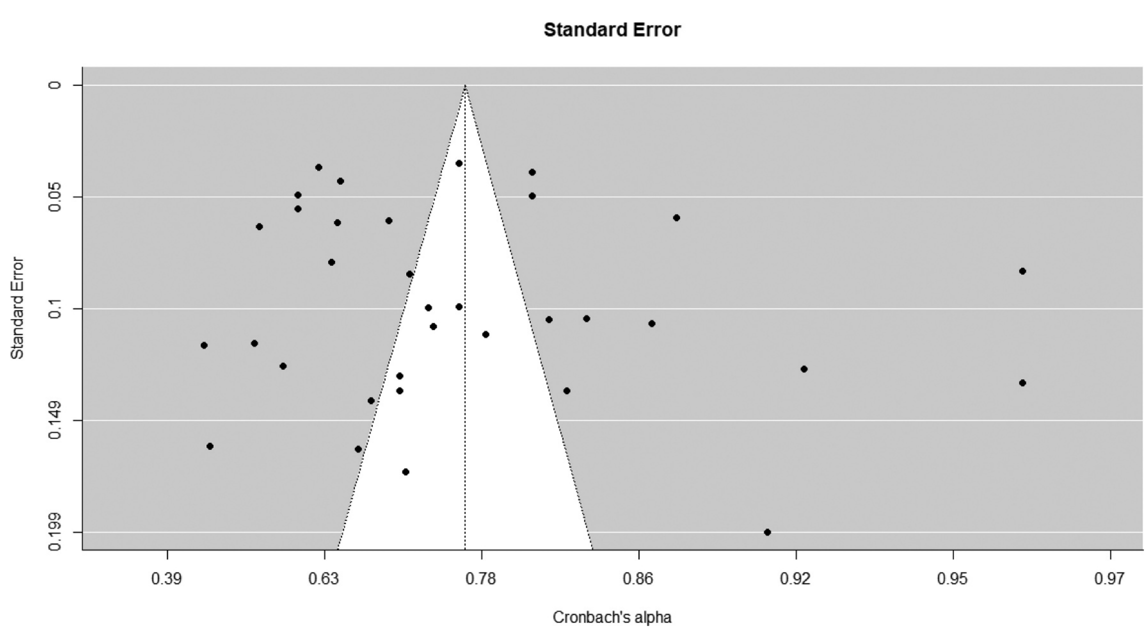

Figure 4. Funnel plot for Cronbach's alpha values.

Table 6. Results of comparing the means of studies that induce and report reliability.

Variable	Inducing Mean(*SD*)	Reporting Mean(*SD*)	*t*	*p*	*d*
Mean age (years)	26.76(9.37) $n_I = 173$	24.15(6.87) $n_R = 23$	1.292	.198	0.396
SD of age (Years)	5.37(3.89) $n_I = 130$	4.81(4.08) $n_R = 22$	0.616	.539	0.143
Female percentage	59.27(16.95) $n_I = 190$	62.12(9.57) $n_R = 26$	−1.269	.210	−0.175

Note. n_I and n_R = sample sizes of inducing and reporting studies, respectively. *t* = *t*-test for comparing two means. *p* = probability level associated to the *t*-test. *d* = standardized mean difference.

analyses. However, the characteristics of the inducing studies were also examined. Several variables, including publication year, publication type (published vs. unpublished), country, population, language, level of education, ethnicity, female proportion, and mean and *SD* of the sample age, were considered to examine the inducing studies.

The percentages were calculated for the categorical variables (e.g., publication type, country, and ethnicity). The publication year of the studies ranged from 2006 to 2021. The studies consisted of journal articles and book chapters (73.30%) or unpublished studies (26.70%), including theses, reports, and proceedings. The studies were conducted in several countries, including the U.S. (50.30%), the UK (7.50%), Canada (5.50%), Italy (5.50%), Netherlands (5.00%), Poland (4.00%), Mixed (4.00%), Germany (3.00%), China (2.50%), Sweden (2.50%), Hungary (2.00%), Taiwan (1.50%), Australia (1.50%), South Africa (1.00%), Austria (0.50%), Hong Kong (0.50%), India (0.50%), Indonesia (0.50%), Ireland (0.50%), Japan (0.50%), Russia (0.50%), and Switzerland (0.50%). The studies were conducted in different continents: Asia (6.00%), Europe (31.50%), North America (55.80%) and others (6.70%). The ethnicities of the samples were Asian (5.70%), Caucasian (39.10%), and mixed (55.20%). The participants of the studies were students (49.00%), individuals from the general population (21.70%), and mixed (29.30%). The educational levels of the student samples were undergraduate (89.60%), graduate (2.10%), high school (5.20%), and mixed (i.e., high school and university students; 3.10%). The English version was administered in 70% of the studies, whereas translated versions (e.g., German, Chinese, Dutch, Italian, and Polish) were used in the rest of the studies (30%).

The characteristics of the inducing studies and reporting studies were compared by means of the *t*-test for the continuous variables (e.g., the means and *SD*s of the sample age, the percentage of females). The results are presented in Table 6. As seen, no statistically significant differences were found for any of the study characteristics.

Discussion

The present study has four objectives: a) estimating the overall reliability coefficient of CAQ scores; b) investigating the impact of certain moderator variables on the reliability coefficients; c) examining the reliability induction rate for the CAQ; and d) comparing the characteristics of the inducing studies with the reporting ones. An RG meta-analysis was conducted to address the first two objectives. To address the last two objectives, the studies were first identified based on estimating, reporting, and omitting the reliability coefficients. Then, the reliability induction rates were estimated, and the characteristics of the inducing studies were examined.

The average alpha value estimated in the present study (.765) is considerably lower than the alpha coefficient reported by Carson et al. (.960; 2005). Nevertheless, the average reliability coefficient is within the acceptable range (O'Rourke et al., 2005). The CAQ is therefore appropriate to be used in exploratory research (Sánchez-Meca et al., 2021). The average reliability coefficient indicates that the standard error of measurement involved in CAQ scores is high and that total scores are not as accurate as needed for high-stakes assessment (i. e., identification of creatively gifted students or important classification of individuals).

The range of the alpha values indicates that the reliability of CAQ scores is sample-dependent. To address this issue, the residual heterogeneity was examined. The variability in the reliability coefficients was high (I^2 = 98.74%). This amount of variance indicates that the reliability of CAQ scores varied significantly across different applications. The residual heterogeneity implies that a reliability coefficient estimated in one sample cannot be generalized to another sample and that certain characteristics of a sample and/or the CAQ impact the reliability of scores. Thus, researchers should not depend on the reliability estimations of other studies and should compute a reliability coefficient for each application.

The variability in the reliability coefficients indicates that some moderator variables affect the reliability of

CAQ scores. Thus, nine continuous and four categorical variables were further examined. Interestingly, none of those variables contributed significantly to the prediction of the overall alpha coefficient. However, this finding is not always necessarily due to low statistical power that originates from invariant alpha values (Onwuegbuzie & Daniel, 2002). When no moderator variable significantly impacts the reliability coefficient, the reliability of scores is considered robust against different applications of the same instrument with respect to those variables (Botella, Suero, & Gambara, 2010). In other words, this type of result may be an indication of the stability of scores on the instrument. Based exclusively on the findings of the present study, we can conclude that the reliability of CAQ scores is stable across sample characteristics and test administration types examined in our study.

Thus, the CAQ can be used with samples that are diverse in age and gender. This point is supported for gender. The literature is filled with evidence against the impact of gender on creative performance (see Baer & Kaufman, 2008; Campos, Lopez, Gonzales, & Perez-Fabello, 2000; Chan et al., 2001; Costa, Terracciano, & Mccrae, 2001; Kaufman, 2006).

The CAQ has been in use for nearly 20 years. Since its development, the CAQ has not been revised and has been regarded as a reliable instrument. The results of the publication year support this conclusion. The CAQ still provides reliable scores.

The findings of the present study suggest that the CAQ can measure creative achievement with a similar level of reliability across cultures, education levels, ethnicities, and countries. Several studies supported this conclusion. Although people from different cultures display different creative performances (Saeki, Fan, & Van Dusen, 2001), findings of previous studies indicate that the view of creativity is quite similar across cultures (see Chen et al., 2002; Kaufman, Baer, & Gentile, 2004; Niu & Sternberg, 2001; Rudowicz, 2003). This point may be an important reason why the reliability of CAQ scores is not significantly different across cultures, countries, and ethnicities.

However, the inconsistency between the results on residual heterogeneity and the results on the moderator variables presents a different picture. It is unexpected to estimate statistically significant heterogeneity in the reliability coefficients but to have all the moderator variables statistically non-significant. There are some potential reasons for obtaining such heterogeneity. The first reason is reliability induction. It should be noted that the total induction rate for the CAQ (85.60%) is higher than the average induction rate mentioned in the literature which is 78.6% (Sánchez-Meca et al., 2015).

Reliability induction directly impacted the number of studies included in this RG meta-analysis. The number of alpha values (N = 34) included in the analyses was enough to obtain interpretable results. However, the low sample size possibly increased the amount of error involved in the results and prevented model development. This situation may have produced nonsignificant results with respect to the moderator variables. In other words, it is likely that the present study lacks the power to detect the impact of the moderator variables due to reliability induction. However, it should be noted that the characteristics of the inducing studies are similar to the reporting studies.

Reliability induction is a highly discouraged practice. The reason is that reliability induction may lead to both overestimation and underestimation of the overall reliability coefficients in RG meta-analyses (Sánchez-Meca et al., 2021). Additionally, when studies do not estimate reliability coefficients, certain moderator variables that should be included in RG analyses and that may significantly impact the average reliability coefficient are overlooked. Finally, reliability induction implies that reliability is an attribute of an instrument and does not change across different administrations of the same instrument (Sánchez-Meca et al., 2021).

There are some reasons why a relatively low number of studies estimated reliability coefficients for CAQ scores. One possible reason is publication bias (Sánchez-Meca et al., 2021). Some researchers may have estimated low reliability coefficients and therefore, may have preferred not to report the coefficients. Nevertheless, it should be noted that our analyses do not indicate publication bias for the CAQ. Another reason may be the assumption that reliability is a property of CAQ scores and does not change across samples. One other potential reason is the dimensionality of the CAQ. It is likely that several researchers chose not to estimate Cronbach's alpha for the CAQ as it is a multi-factor instrument – note that previous studies found that multi-factor (two-factor or three-factor) models had better model-data fit than the single-factor model (e.g., Balta et al., 2019; Carson et al., 2005). In nineteen percent of studies that committed reliability induction by report and 8.9% of studies that committed reliability induction by omission, researchers explicitly stated that the CAQ is a multi-factor instrument. The multifactor structure of the CAQ was not mentioned in the rest of the inducing studies, but it is possible that the researchers may have preferred not to estimate Cronbach's alpha due to the dimensionality of the CAQ.

The second reason for obtaining residual heterogeneity is related to certain moderator variables that were not included in the analyses of the present study. Certain descriptive statistics were reported by the previous studies and were included in the present study as moderators. However, some other moderator variables were not reported by the overwhelming majority of studies and were not analyzed.

One of those variables is the number of factors impacting the CAQ scores. Previous studies (Balta et al., 2019; Carson et al., 2005) identified multifactor structures for the CAQ, but a considerable number of studies that reported reliability coefficients treated the CAQ as a single-factor instrument (e.g., Hezel & Hooley, 2014; Ruiter & Johnson, 2015). It is possible that the CAQ behaved as a multi-factor instrument in those studies. This multi-factor structure may have impacted the reliability of scores across studies. However, the number of studies that included data on the dimensionality of the CAQ was not enough to designate the number of factors as a moderator variable.

The issue of whether creativity is a domain-specific or domain-general skill is at the center of creativity research. Research on creative products supports the hypothesis that creativity is domain-specific, whereas research on creative persons indicates that creativity is domain-general (Silvia, Kaufman, & Pretz, 2009). Because the CAQ assesses one's creative achievement derived from his or her creative products, it is expected to detect domain-specificity of creativity in the CAQ and to identify a multi-factor structure for the instrument. In other words, as suggested by Carson et al. (2005), creative achievement in the arts should be a separate phenomenon from creative achievement in science.

In fact, factor analysis of the CAQ shows that the art domains are impacted by one factor and the science domains are affected by another factor (Balta et al., 2019; Carson et al., 2005). A study conducted by Silvia et al. (2009) provided additional evidence for the domain-specificity of the CAQ. Silvia et al. examined the latent classes in the CAQ and identified three classes. This particular study showed that some participants were creative in the visual arts, while some participants were creative in the performing arts, including music, creative writing, and theater/film. The third class was composed of participants who did not exhibit high creative achievement in any area.

One of the assumptions regarding Cronbach's alpha is unidimensionality (Tavakol & Dennick, 2011). The alpha value fluctuates when an instrument has a multi-factor structure (Tavakol & Dennick, 2011). Thus, Cronbach's alpha should be estimated if unidimensionality and tau-equivalency are met (Tavakol & Dennick, 2011). When we look at the structure of the CAQ, it seems to be a multidimensional instrument (see Carson et al., 2005). In fact, factor analysis studies indicate that the CAQ is multidimensional. Additionally, the items on the CAQ are possibly not tau-equivalent (for details, see Balta et al., 2019). Considering the unique nature of the CAQ, Cronbach's alpha does not seem to be the ideal option for estimating the reliability of scores. There are several alternatives to Cronbach's alpha, such as the composite reliability (Raykov, 1997), Omega (McDonald, 1970), coefficient H (Hancock & Mueller, 2001), and greatest lower bound (Jackson & Agunwamba, 1977). Another choice is to rely on temporal stability. However, estimating a test-retest reliability coefficient may not be realistic for cross-sectional studies.

Another variable that may have impacted the reliability coefficients is the group characteristics. The majority of previous studies did not categorize the participants based on their creative achievement levels. Instead, CAQ scores were used for different purposes, such as factor analyzing the data (Balta et al., 2019), correlating CAQ scores with other variables (Kéri, 2011), or comparing the CAQ scores of different subgroups (Duan, Wang, Hu, & Kounios, 2020). Few studies mentioned the creativity levels of the participants (see Diedrich et al., 2018; Silvia & Nusbaum, 2012; Verger, Shankland, & Sudres, 2021). For instance, Diedrich et al. analyzed the CAQ data obtained from "little-c samples" (2018, p. 3). On the other hand, Verger et al. stated that their "study specifically targeted individuals with presumed little-c (everyday) creativity levels" although some participants might display a "Pro-c level of creativity" (2021, p. 5). Finally, Silvia and Nusbaum (2012) categorized the participants as creative or non-creative based on their major (i.e., "people with minors or with more than one major, having any creative arts major or minor was enough to be scored as creative" p. 34). It is possible that the group characteristics impact the reliability of CAQ scores. However, the number of studies that categorized the participants as creative experts and creative achievers or as creative and non-creative was not enough to designate the group characteristics as a moderator variable.

One may argue that different levels of creative achievement can be captured by the mean score. However, this is possibly not the case for the studies included in the analyses. The reason is that the level of creative achievement does not necessarily increase as the total CAQ score goes up. For instance, an individual at a high creativity level may have displayed one nationally recognized achievement, and therefore, he or she earns 7 points for an item with an asterisk; whereas another individual at the same level may have displayed four nationally recognized achievements, and therefore, he or she earns 28 points for the same item. In other words, different individuals who are at the same level of creative achievement may have considerably different total scores. Additionally, an individual who is creative in more than one domain may receive the same score as an individual who is at a higher level in one domain. This means that two individuals with the same score may be at different levels of creative achievement.

In sum, the moderators included in the analyses failed to reach statistical significance, but this does not necessarily mean that the reliability of CAQ scores is robust against the effects of different variables. The reason is that the variability in the overall reliability coefficient was high. Some variables which were not included in the analyses of the present study can explain the residual in the reliability coefficients.

There are certain limitations of the present study. Although different reliability coefficients were available, we included only Cronbach's alpha coefficient in this RG analysis. Another limitation is the number of studies included in the analyses. This limitation is directly related to reliability induction. Finally, the incomplete reporting of the moderator variables is another limitation.

The present study shows that estimating the reliability coefficient is an important responsibility of researchers. Thus, we recommend all researchers to compute reliability coefficients and to report them even if the coefficients are low. Additionally, we urge researchers to identify the characteristics of their samples and report important descriptive statistics thoroughly. Cronbach's alpha may not be the ideal method for estimating the reliability of CAQ scores (Silvia et al., 2012). We recommend researchers to use other methods for internal consistency.

Future research about the reliability generalization of the CAQ is needed. Researchers should consider using other transformation methods and multivariate models. Additionally, if future studies start reporting different types of reliability coefficients and descriptive statistics, future research should focus on other reliability indicators and examine the impact of other moderator variables.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Süreyya Yörük http://orcid.org/0000-0001-8181-2614
Sedat Sen http://orcid.org/0000-0001-6962-4960

References

Note: References with an asterisk (*) indicate studies included to reliability generalization meta-analysis

Ahmetoglu, G. (2014). *The entrepreneurial personality: A new framework and construct for entrepreneurship research and practice* [Unpublished doctoral dissertation]. University of London.

*Ahmetoglu, G., Harding, X., Akhtar, R, & Premuzic, T. (2015). Predictors of creative achievement: Assessing the impact of entrepreneurial potential, perfectionism, and employee engagement. *Creativity Research Journal*, *27*(2), 198–205. doi:10.1080/10400419.2015.1030293

Badibanga, J. A. (2013). *Cultural values and creativity: A two country comparison between the United States and South Africa* [Unpublished master's thesis]. University of North Florida.

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

Balta, M., Subotić, S., & Pavićević, M. D. (2019). *Factor structure and convergent validity of the creative achievement questionnaire (CAQ)*. Proceedings of the XXV Scientific Conference, Belgrade, Serbia. 92–94.

Barron, F. (1955). The disposition toward originality. *Journal of Abnormal and Social Psychology*, *51*(3), 478–485. doi:10.1037/h0048073

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, *50*(4), 1088–1101. doi:10.2307/2533446

Bojner-Horwitz, E., Lennartsson, A. K., Theorell, T. P., & Ullén, F. (2015). Engagement in dance is associated with emotional competence in interplay with others. *Frontiers in Psychology*, *6*, 1096. doi:10.3389/fpsyg.2015.01096

Bonett, D. G. (2002). Sample size requirements for testing and estimating coefficient alpha. *Journal of Educational and Behavioral Statistics*, *27*(4), 335–340. doi:10.3102/10769986027004335

Botella, J., Suero, M., & Gambara, H. (2010). Psychometric inferences from a meta-analysis of reliability and internal consistency coefficients. *Psychological Methods*, *4*(4), 386–397. doi:10.1037/a0019626

Campos, A., Lopez, A., Gonzales, M. A., & Perez-Fabello, M. J. (2000). Aspects of creativity affected by imaging capacity. *North American Journal of Psychology*, *2*, 313–321.

Carruthers, L. (2016). *Creativity and attention: A multi-method investigation* [Unpublished doctoral dissertation]. Edinburgh Napier University.

Carson, S. H. (2001). *Demons and muses: An exploration of cognitive features and vulnerability to psychosis in creative individuals* [Unpublished doctoral dissertation]. Harvard University.

*Carson, S. H., Peterson, J. B., & Higgins, D. M. (2005). Reliability, validity, and factor structure of the creative achievement questionnaire. *Creativity Research Journal*, *17*(1), 37–50. doi:10.1207/s15326934crj1701_4

Chai, J., & Fan, -K.-K. (2016). Mobile inverted constructivism: Education of interaction technology in social media. *Eurasia Journal of Mathematics, Science and Technology Education*, *12*(5), 1425–1442. doi:10.12973/eurasia.2016.1522a

Chan, D. W., Cheung, P. C., Lau, S., Wu, W. Y., Kwong, J. M., & Li, W. L. (2001). Assessing ideational fluency in primary students in Hong Kong. *Creativity Research Journal*, *13*(3–4), 359–365. doi:10.1207/S15326934CRJ1334_13

Chen, C., Kasof, J., Himsel, A. J., Greenberger, E., Dong, Q., & Xue, G. (2002). Creativity in drawings of geometric shapes: A cross-cultural examination with the consensual assessment technique. *Journal of Cross-Cultural Psychology*, *33*(2), 171–187. doi:10.1177/0022022102033002004

*Conover, K. J. (2015). *Ableist microaggressions scale: Development, validation, and relationship with social support* [Unpublished doctoral dissertation]. Santa Barbara: University of California.

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

de Manzano, Ö., & Ullén, F. (2018). Genetic and environmental influences on the phenotypic associations between intelligence, personality, and creative achievement in the arts and sciences. *Intelligence*, *69*, 123–133. doi:10.1016/j.intell.2018.05.004

*Deng, L., Wang, L., & Zhao, Y. (2016). How creativity was affected by environmental factors and individual characteristics: A cross-cultural comparison perspective. *Creativity Research Journal*, *28*(3), 357–366. doi:10.1080/10400419.2016.1195615

*Diedrich, J., Jauk, E., Silvia, P. J., Gredlein, J. M., Neubauer, A. C., & Benedek, M. (2018). Assessment of real-life creativity: The inventory of creative activities and achievements (ICAA). *Psychology of Aesthetics, Creativity, and the Arts*, *12*(3), 304–316. doi:10.1037/aca0000137

*Dostál, D., Plháková, A., & Záškodná, T. (2015). Domain-specific creativity in relation to the level of empathy and systemizing. *Journal of Creative Behavior*, *51*(3), 225–239. doi:10.1002/jocb.103

*Duan, H., Wang, X., Hu, W., & Kounios, J. (2020). Effects of acute stress on divergent and convergent problem-solving. *Thinking & Reasoning*, *26*(1), 68–86. doi:10.1080/13546783.2019.1572539

Du, Y., Xie, L., Zhong, J. A., Zou, H., Law, R., & Yan, X. (2019). Creativity fostering teacher behavior on student creative achievement: Mediation of intrinsic motivation and moderation of openness to experience. *School Psychology International*, *40*(5), 525–542. doi:10.1177/0143034319868271

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*(7109), 629–634. doi:10.1136/bmj.315.7109.629

*Falavarjani, M. F., & Yeh, C. J. (2018). The impact of acculturation identification and acculturative stress on creativity among Iranian immigrants living in Malaysia. *Journal of Ethnic and Migration Studies*, *44*(13), 2219–2239. doi:10.1080/1369183X.2017.1366301

*Forgeard, M. J. C., & Benson, L. (2019). Extracurricular involvement and psychological adjustment in the transition from adolescence to emerging adulthood: The role of mastery and creative self-efficacy. *Applied Developmental Science*, *23*(1), 41–58. doi:10.1080/10888691.2017.1288124

Form, S., Aue, J., & Kaernbach, C. (2019). Judging popular novels as creative products: Which creative attributes contribute to their success? *Creativity Theories, Research, and Applications*, *6*(1), 21–41. doi:10.1515/ctra-2019-0002

*Grohman, M. G., Ivcevic, Z., Silvia, P., & Kaufman, S. B. (2017). The role of passion and persistence in creativity. *Psychology of Aesthetics, Creativity, and the Arts*, *11*(4), 376 385. doi:10.1037/aca0000121

Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. du Toit, & D. Soerbom (Eds.), *Structural equation modeling: Present and future—A festschrift in honor of Karl Jöreskog* (pp. 195–216). Lincolnwood, IL: Scientific Software International.

*Han, W., Zhang, M., Feng, X., Gong, G., Peng, K., & Zhang, D. (2018). Genetic influences on creativity: An exploration of convergent and divergent thinking. *Peer J*, *6*, 1–18. doi:10.7717/peerj.5403

*Hezel, D. M., & Hooley, J. M. (2014). Creativity, personality, and hoarding behavior. *Psychiatry Research*, *220*(1–2), 322–327. doi:10.1016/j.psychres.2014.07.037

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*(7414), 557–560. doi:10.1136/bmj.327.7414.557

Hocevar, D. (1979, April). *The development of the creative behavior inventory*. Paper presented at the annual meeting of the Rocky Mountain Psychological Association, Albuquerque, NM, United States.

Jackson, P. H., & Agunwamba, C. C. (1977). Lower bounds for the reliability of the total score on a test composed of nonhomogeneous items. I: Algebraic lower bounds. *Psychometrika*, *42*(4), 567–578. doi:10.1007/BF02295979

*Jacobs, M., Huijts, T., van Broekhoven, K., Straetmans, S., Meng, C., & van der Velden, R. (2021). Effects of participation in excellence programmes on cognitive and non-cognitive skills among higher education students in the Netherlands. *European Journal of Higher Education*, *11*(1), 44–63. doi:10.1080/21568235.2020.1850315

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

*Kaufman, S. B. (2009). *Beyond general intelligence: The dual-process theory of human intelligence* [Unpublished doctoral dissertation]. Yale University.

Kaufman, J. C., Baer, J., & Gentile, C. A. (2004). Differences in gender and ethnicity as measured by ratings of three writing tasks. *Journal of Creative Behavior*, *39*(1), 56–69. doi:10.1002/j.2162-6057.2004.tb01231.x

Kaufman, J. C., & Beghetto, R. A. (2009). Beyond big and little: The four c model of creativity. *Review of General Psychology*, *13*(1), 1–12. doi:10.1037/a0013688

*Kéri, S. (2011). Solitary minds and social capital: Latent inhibition, general intellectual functions and social network size predict creative achievements. *Psychology of Aesthetics, Creativity, and the Arts*, *5*(3), 215–221. doi:10.1037/a0022000

Kuckelkorn, K. L., de Manzano, Ö., & Ullén, F. (2021). Musical expertise and personality-differences related to occupational choice and instrument categories. *Personality and Individual Differences*, *173*, 110573. doi:10.1016/j.paid.2020.110573

*Lebuda, I., Zielińska, A., & Karwowski, M. (2021). On surface and core predictors of real-life creativity. *Thinking Skills and Creativity*, *42*, 1–15. doi:10.1016/j.tsc.2021.100973

*Loh, K. K., & Lim, S. W. (2020). Positive associations between media multitasking and creativity. *Computers in Human Behavior Reports*, *1*, 1–14. doi:10.1016/j.chbr.2020.100015

Lu, C. (2017). Interactive effects of environmental experience and innovative cognitive style on student creativity in product design. *International Journal of Technology and Design Education*, *27*(4), 577–594. doi:10.1007/s10798-016-9368-x

*Lucas, G. J. M., van der Wijst, A., Curşeu, L. P., & Looman, W. L. (2013). An evaluation of alternative ways of computing the creativity quotient in a design school sample. *Creativity Research Journal*, *25*(3), 348–355. doi:10.1080/10400419.2013.813811

Ludwig, A. (1992). The creative achievement scale. *Creativity Research Journal*, *5*(2), 109–124. doi:10.1080/10400419209534427

McDonald, R. P. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology*, *23*(1), 1–21. doi:10.1111/j.2044-8317.1970.tb00432.x

*Mendes, N., Oligschläger, S., Lauckner, M. E., Golchert, J., Huntenburg, J. M., Falkiewicz, M., … Margulies, D. S. (2019). A functional connectome phenotyping dataset including cognitive state and personality measures. *Scientific Data*, *6*(1), 180307. doi:10.1038/sdata.2018.307

Mosing, M. A., Verweij, K. J., Madison, G., Pedersen, N. L., Zietsch, B. P., & Ullén, F. (2015). Did sexual selection shape human music? Testing predictions from the sexual selection hypothesis of music evolution using a large genetically informative sample of over 10,000 twins. *Evolution and Human Behavior*, *36*(5), 359–366. doi:10.1016/j.evolhumbehav.2015.02.004

Mumford, M. D. (2003). Taking stock in taking stock. *Creativity Research Journal*, *15*, 147–151. doi:10.1080/10400419.2003.9651408

Niu, W., & Sternberg, R. J. (2001). Cultural influences on artistic creativity and its evaluation. *International Journal of Psychology*, *36*(4), 225–241. doi:10.1080/00207590143000036

Ofili, P. (2011). *Do multicultural experiences and biculturalism promote creativity in international students?* [Unpublished master's thesis]. University of Mississippi.

*Onwe, N. S. (2019). *Analyzing the connection between work engagement factors and creativity* [Unpublished master's thesis]. Austin Peay State University.

Onwuegbuzie, A. J., & Daniel, L. G. (2002). A Framework for reporting and interpreting internal consistency reliability estimates. *Measurement and Evaluation in Counseling and Development*, *35*(2), 89–103. doi:10.1080/07481756.2002.12069052

O'quin, K., & Besemer, S. P. (1999). Creative products. In M. Runco & S. Pritzker (Eds.), *Encyclopedia of creativity* (pp. 413–422). London: Academic Press.

O'Rourke, N., Hatcher, L., & Stepanski, E. J. (2005). *A step-by-step approach to using SAS for univariate & multivariate statistics*. Cary, NC, United States: SAS Institute.

*Patil, A. U., Madathil, D., & Huang, C. M. (2021). Healthy aging alters the functional connectivity of creative cognition in the default mode network and cerebellar network. *Frontiers in Aging Neuroscience*. *13*. Article 607988. doi:10.3389/fnagi.2021.607988.

Polner, B., Nagy, H., Takáts, A., & Kéri, S. (2015). Kiss of the muse for the chosen ones: De novo schizotypal traits and lifetime creative achievement are related to changes in divergent thinking during dopaminergic therapy in Parkinson's disease. *Psychology of Aesthetics, Creativity, and the Arts*, *9*(3), 328–339. doi:10.1037/a0039303

Rahmani, F., Moghaddam, H. S., & Aarabi, M. H. (2020). Intact microstructure of the right corticostriatal pathway predicts creative ability in healthy adults. *Brain and Behavior*, *10*(12), 1–10. doi:10.1002/brb3.1895

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, *21*(2), 173–184. doi:10.1177/01466216970212006

Richards, R., Kinney, D., Benet, M., & Merzel, A. (1988). Assessing everyday creativity: Characteristics of the lifetime creativity scales and validation with three large samples. *Journal of Personality and Social Psychology*, *54*(3), 476–485. doi:10.1037/0022-3514.54.3.476

*Rodet, C. S. (2021). Experiments on creativity and work design. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3647303

Rosenberg, M. S. (2005). The file-drawer problem revisited: A general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution*, *59*(2), 464–468. doi:10.1111/j.0014-3820.2005.tb01004.x

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. doi:10.1037/0033-2909.86.3.638

*Rudowicz, E. (2003). Creativity and culture: A two way interaction. *Scandinavian Journal of Educational Research*, *47*(3), 273–290. doi:10.1080/00313830308602

Ruiter, M., & Johnson, S. L. (2015). Mania risk and creativity: A multi-method study of the role of motivation. *Journal of Affective Disorders*, *170*, 52–58. doi:10.1016/j.jad.2014.08.049

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

Saeki, N., Fan, X., & Van Dusen, L. (2001). A comparative study of creative thionine of American and Japanese college students. *Journal of Creative Behavior*, *35*(1), 24–36. doi:10.1002/j.2162-6057.2001.tb01219.x

Salvi, C., Costantini, G., Pace, A., & Palmiero, M. (2020). Validation of the Italian remote associate test. *The Journal of Creative Behavior*, *54*(1), 62–74. doi:10.1002/jocb.345

Sánchez-Meca, J., Marín-Martínez, F., López-López, J. A., Núñez-Núñez, R. M., Rubio-Aparicio, M., López-García, J. J., … López-Nicolás, R. (2021). Improving the reporting quality of reliability generalization meta-analyses: The REGEMA checklist. *Research Synthesis Methods*, *12*(4), 516–536. doi:10.1002/jrsm.1487

Sánchez-Meca, J., Rubio-Aparicio, M., López-Pina, J., Núñez-Núñez, R. M., & Marín-Martínez, F. (2015July). The phenomenon of reliability induction in the social and health sciences. In *Paper presented at the XIV Congress of Methodology of the Social and Health Sciences*. Palma de Mallorca, Spain.

Sen, S. (2022). A reliability generalization meta-analysis of Runco ideational behavior scale. *Creativity Research Journal*, *34*(2), 178–194. doi:10.1080/10400419.2021.1960719

*Shelton, K. D. (2010). *Autonomy and organizational justice as moderators of the relationships among creativity, openness to experience, and organizational misbehavior* [Unpublished doctoral dissertation]. Walden University.

Silvia, P. J., Kaufman, J. C., & Pretz, J. E. (2009). Is creativity domain-specific? Latent class models of creative accomplishments and creative self-descriptions. *Psychology of Aesthetics, Creativity, and the Arts*, *3*(3), 139–148. doi:10.1037/a0014940

Silvia, P. J., Kaufman, J. C., Reiter-Palmon, R., & Wigert, B. (2011). Cantankerous creativity: Honesty–humility, agreeableness, and the HEXACO structure of creative achievement. *Personality and Individual Differences*, *51*(5), 687–689. doi:10.1016/j.paid.2011.06.011

*Silvia, P. J., & Nusbaum, E. C. (2012). What's your major? College majors as markers of creativity. *International Journal of Creativity and Problem-Solving*, *22*(2), 31–43.

*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. doi:10.1037/a0024071

*Specker, E., Forster, M., Brinkmann, H., Boddy, J., Pelowski, M., Rosenberg, R., & Leder, H. (2020). The Vienna art interest and art knowledge questionnaire (VAIAK): A unified and validated measure of art interest and art knowledge. *Psychology of Aesthetics, Creativity, and the Arts*, *14*(2), 172–185. doi:10.1037/aca0000205

Stein, M. I. (1953). Creativity and culture. *Journal of Psychology*, *36*(2), 311–322. doi:10.1080/00223980.1953.9712897

*Sutu, A., Serrano, S., Schultz, L. H., Jackson, J. J., & Damian, R. I. (2019). Creating through deviancy or adjustment? The link between personality profile normativeness and creativity. *European Journal of Personality*, *33*(5), 565–588. doi:10.1002/per.2215

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. doi:10.5116/ijme.4dfb.8dfd

Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement*, *58*(1), 6–20. doi:10.1177/0013164498058001002

Vellante, M., Zucca, G., Preti, A., Sisti, D., Rocchi, M. B., Akiskal, K. K., & Akiskal, H. S. (2011). Creativity and affective temperaments in non-clinical professional artists: An empirical psychometric investigation. *Journal of Affective Disorders*, *135*(1–3), 28–36. doi:10.1016/j.jad.2011.06.062

*Verger, N. B., Shankland, R., & Sudres, J. L. (2021). High artistic achievements and low emotion dysregulation: The moderating and mediating role of self-compassion. *Creativity Research Journal*, *34*(1), 68–84. doi:10.1080/10400419.2021.1962104

Viechtbauer, W. (2010). *Metafor meta-analysis package for R, Version 1.4-0* [Computer software, Accessed 20 January 2022]. http://CRAN.R-project.org/packag=metafor

*von Stumm, S., & Scott, H. (2019). Imagination links with schizotypal beliefs, not with creativity or learning. *British Journal of Psychology*, *110*(4), 707–726. doi:10.1111/bjop.12369

Wang, C., Ho, H., Cheng, C., & Cheng, Y. (2014). Application of the Rasch model to the measurement of creativity: The creative achievement questionnaire. *Creativity Research Journal*, *26*(1), 62–71. doi:10.1080/10400419.2013.843347

Wechsler, D. (1981). *WAIS–R manual: Wechsler adult intelligence scale–revised*. San Antonio, TX: Psychological Corporation.

*Willard, R. M. (2017). *Rumination and depressive psychopathology as predictors of creative achievement* [Unpublished master's thesis]. Illinois State University.

Zhanqi, W., & Xingli, Z. (2020). The validity and reliability of the Chinese version of the creative achievement questionnaire. *Studies of Psychology and Behavior*, *18*(3), 390–397.

*Zhu, W., Shang, S., Jiang, W., Pei, M., & Su, Y. (2019). Convergent thinking moderates the relationship between divergent thinking and scientific creativity. *Creativity Research Journal*, *31*(3), 320–328. doi:10.1080/10400419.2019.1641685