

A Reliability Generalization Meta-Analysis of the Kaufman Domains of Creativity Scale

ABSTRACT

The Kaufman Domains of Creativity Scale (K-DOCS) is a self-reported rating scale that measures creative behaviors in five areas. Despite the vast amount of research on the scale, the internal consistency reliability of K-DOCS scores have not been examined. Specifically, there is no study on the overall reliability coefficients, the variation in the reliability of scores, and reliability induction. In the present study, reliability generalization meta-analyses were conducted to address these issues. The sample consisted of 56 studies that produced 60 Cronbach's alpha coefficients in total. The pooled alpha values were estimated to be .904 (total scale), .825 (Self/Everyday subscale), .858 (Scholarly subscale), .887 (Performance subscale), .867 (Scientific subscale), and .861 (Artistic subscale). The moderator analyses showed that the reliability estimates of K-DOCS total scores and Self/Everyday subscale scores did not differ with respect to any of the moderator variables. On the other hand, certain variables affected the alpha coefficients for Scholarly (educational level, *SD* of the age, and mean age), Performance (continent, educational level, ethnicity, Caucasian percentage, *SD* of the age, and mean age), Scientific (language, test version, continent, country, ethnicity, *SD* of the age, and mean age), and Artistic (continent, language, country, mean age, and total mean score) subscale scores. Finally, the reliability induction rate was 39.62%, and there were no statistically significant differences between the inducing and reporting studies with respect to the continuous variables (mean of the total score, means and *SDs* of the sample age, and percentages of female and Caucasian). Our findings indicate that the K-DOCS provides highly reliable scores. However, certain variables contribute to systematic errors in Scholarly, Performance, Scientific, and Artistic subscale scores. Hence, scores on these subscales should be interpreted with caution. Due to a high amount of variation in the reliability coefficients, reliability induction is not advised for the K-DOCS.

Keywords: kaufman domains of creativity scale, K-DOCS, reliability generalization meta-analysis, reliability induction.

Creativity plays a pivotal role across diverse domains, spanning from the arts and sciences to business and education. For instance, in artistic domains, creativity leads to artwork that evokes emotions, challenges conventions, and captures the essence of the human experience. In business, creativity enables organizations to adapt, differentiate, and seize opportunities in markets. At its core, creativity empowers individuals to perceive the world through fresh lenses, envision new horizons, and generate original as well as useful ideas, solutions, perspectives, and products (Funke, 2009; Mumford, 2003; Runco, 2007; Runco & Jaeger, 2012; Starko, 2014; Stein, 1953).

Creativity manifests itself in two forms: creative potential and creative performance (Guilford, 1966; Hinton, 1968). Creative potential refers to an individual's capacity to generate novel and functional ideas or products (Guilford, 1966; Runco et al., 2001). On the other hand, creative performance pertains to the tangible outcomes (e.g., a song and a novel) and practical manifestation of creative potential (Guilford, 1966). Therefore, the concept of creativity is shaped by the interplay between creative potential and creative performance.

DOMAIN-GENERALITY AND DOMAIN-SPECIFICITY OF CREATIVITY

There are three different views on the nature of creativity. Some researchers argue for the domain-general view, which suggests that individuals must possess domain-general skills to produce creative work in any domain (e.g., Chen et al., 2006; Plucker, 1998, 1999, 2004; Simonton, 2017). Research on the

characteristics of creative people provides supporting evidence for this standpoint (Silvia et al., 2009). Some other researchers propose that creativity is domain-specific (e.g., Baer, 1998, 2012, 2015; Boccia et al., 2015; Dow & Mayer, 2004). In other words, each domain requires a different set of thinking skills, and creative thinking skills in one domain cannot be transferred to another. Research on creative performance offers evidence that supports this perspective (Silvia et al., 2009). Finally, there are researchers who hypothesize that creativity entails both domain-general and domain-specific requirements (e.g., Baer & Kaufman, 2005; Plucker & Beghetto, 2004; Sternberg, 2009).

An example of this third view is the Amusement Park Theory (Baer & Kaufman, 2005), which presents a four-level model. As the levels move forward from the first to the last, each subsequent level becomes more domain-specific. According to the model, a person needs to possess both domain-general skills (e.g., general intellectual ability and desire to generate work) and domain-specific skills (e.g., style of thinking) in thematic areas (e.g., the arts and science), domains (e.g., music and literature domains in the arts), and micro-domains (e.g., piano or guitar in the music domain) to produce creative work.

The Amusement Park Theory highlights that researchers need to identify specific factors in a particular area, domain, or micro-domain to develop instruments that can reliably, validly, and fairly measure creative potential or creative performance. However, there are numerous areas (e.g., the arts and sports), domains (e.g., music and painting in the arts), and micro-domains (e.g., piano and guitar in music) to consider (Baer & Kaufman, 2005). Additionally, there are various ways to measure creativity, such as employing divergent thinking tasks, observing behaviors, and rating products. Hence, the development of reliable, valid, and fair instruments to measure creativity across different domains poses considerable challenges.

ASSESSMENT OF CREATIVITY

Different instruments or methods are available for measuring creativity in various domains. Instruments or methods that intend to assess individuals' achievements (e.g., the Creative Achievement Questionnaire; Carson et al., 2005) and products (e.g., the Consensual Assessment Technique; Amabile, 1982) are two examples. The Creative Achievement Questionnaire (Carson et al., 2005) asks respondents to report their past achievements in 10 domains and measures creative performance. The Consensual Assessment Technique (Amabile, 1982) makes it possible to evaluate creative products based on expert judgment.

Another approach involves the use of instruments that focus on everyday creative behaviors in certain domains. The Creative Behavior Inventory (Hocevar, 1979) and the Kaufman Domains of Creativity Scale (K-DOCS; Kaufman, 2012) are examples of this final approach. The K-DOCS differs from the Creative Behavior Inventory in that the former requires respondents to rate their creativity level on certain behaviors (e.g., "Making up rhymes;" Kaufman, 2012, p. 307) compared to other people, while the latter asks respondents to report whether they displayed certain behaviors (e.g., "Designed a game;" Hocevar, 1979, p. 15).

Although both are respected instruments, considerably more researchers prefer to use the K-DOCS to measure creative behavior. As of August 21 2023, the number of studies that cited the K-DOCS ($n = 467$) is much higher than the number of studies that cited the Creative Behavior Inventory ($n = 248$) based on the results on Google Scholar. This is possibly due to the psychometric qualities (see Brauer et al., 2022; Kapoor et al., 2021; Kaufman, 2012; Miroshnik et al., 2022), easy-to-use structure, and recentness of the K-DOCS. Additionally, the K-DOCS stands out among other instruments, as it was developed to measure everyday creative behaviors from a layperson's perspective in a short amount of time (Kaufman, 2012). Furthermore, even if respondents have not exhibited a certain behavior covered in a particular item, the K-DOCS allows respondents to rate themselves based on behaviors similar to those targeted by the item (Kaufman, 2012).

Hence, the present study examines the K-DOCS. Specifically, this study investigates the internal consistency reliability of K-DOCS scores. We followed the guidelines explained in the REGEMA (reliability generalization meta-analysis) checklist for conducting the study (see Sánchez-Meca et al., 2021).

THE KAUFMAN DOMAINS OF CREATIVITY SCALE

The K-DOCS is a self-reported rating scale that is intended to measure creative performance (i.e., behavior) in five areas (Kaufman, 2012). Those areas are as follows: Self/Everyday (items 1–11), Scholarly (items 12–22), Performance (items 23–32), Scientific (items 33–41), and Artistic (items 42–50). Each area covers certain domains. For instance, the Performance area (i.e., subscale) encompasses writing and music (Kaufman, 2012).

When developing the scale, Kaufman (2012) created an item pool composed of 94 items based on the Creativity Domain Questionnaire (Kaufman et al., 2009). Those 94 items were analyzed through exploratory

factor analysis (EFA) to determine the best items and identify the factors. The sample was composed of 2,318 undergraduate students in the United States. Fifty items with factor loadings over .45 were retained after the analyses. Those 50 items were loaded on five factors, each of which had an eigenvalue greater than 2.00. Finally, considering the target behaviors of the items, those five factors were labeled as the “Self/Everyday Creativity,” “Scholarly Creativity,” “Performance Creativity,” “Mechanical/Scientific Creativity,” and “Artistic Creativity” (Kaufman, 2012, pp. 300–302).

The retained 50 items constitute the original version of the K-DOCS. Each item is on a 5-point scale, ranging from “*much less creative*” to “*much more creative*” (Kaufman, 2012, p. 307). The lowest response option is given 1 point, while the highest option is given 5 points. There is no hierarchy among the items within a subscale.

The Self/Everyday subscale focuses on intrapersonal (e.g., “Understanding how to make myself happy;” Kaufman, 2012, p. 307) and interpersonal (e.g., “Helping other people cope with a difficult situation;” Kaufman, 2012, p. 307) creative behaviors. The Scholarly subscale is related to the use of knowledge in certain domains that involve verbal or written communication (e.g., “Debating a controversial topic from my own perspective;” Kaufman, 2012, p. 307). The Performance subscale focuses on behaviors that require some level of psychomotor activity in certain domains, such as music and writing (e.g., “Learning how to play a musical instrument;” Kaufman, 2012, p. 307). The Scientific subscale encompasses certain behaviors in science, engineering, and math (e.g., “Writing a computer program;” Kaufman, 2012, p. 307). Finally, the Artistic subscale covers art-related behaviors in certain domains such as painting, photography, and sculpture (e.g., “Making a sculpture or piece of pottery;” Kaufman, 2012, p. 307).

The items require respondents to compare themselves to other people regarding certain behaviors and to rate themselves in those behaviors (see Kaufman, 2012 for details). The following instructions are given to the respondents:

Compared to people of approximately your age and life experience, how *creative* would you rate yourself for each of the following acts? For acts that you have not specifically done, estimate your creative potential based on your performance on similar tasks. (Kaufman, 2012, p. 300)

Kaufman (2012) did not explain the target populations of the K-DOCS. However, he stated that the original study “allow[ed] only a certain amount of extrapolation to the general population” (Kaufman, 2012, p. 304). Thus, it can be inferred that the target populations consist of individuals from the general population, regardless of age and field of interest or expertise.

Translated versions of the K-DOCS are available in different languages, such as Chinese (Wang et al., 2020), Nigerian (Awofala & Fatade, 2015), Russian (Miroshnik et al., 2022), and Turkish (Şahin, 2016a). To our knowledge, the Chinese (Zhao et al., 2017), Croatian (Benić, 2021), Kazakh (Altınay et al., 2022), Nigerian (Awofala & Fatade, 2015), Polish (McKay et al., 2017), and Turkish (Şahin, 2016a) versions of the scale were standardized in languages other than English. The other administrations were performed with mere translations.

Different versions of the K-DOCS are available. In addition to the full version with 50 items, there are validated 42-item (Şahin, 2016a) and 20-item (Tan et al., 2021) versions. Note that the former is in Turkish and the latter is in English. The short versions also measure creative behaviors in five areas. Apart from these versions, some researchers use modified versions of the scale—note that those modifications are not based on any statistical analysis (e.g., Ashraf et al., 2019; Jung et al., 2021).

Psychometric properties of the K-DOCS

Validity of the original version

The evidence for the structural and generalizability aspects of the validity of the K-DOCS is satisfactory. The five-factor model identified by Kaufman (2012) and measurement invariance were supported in previous studies. Kapoor et al. (2021) analyzed data obtained from 22,013 respondents to examine the structural and generalizability aspects of validity. Note that the data were collected from university students in the United States in previous studies. A confirmatory factor analysis (CFA) and measurement invariance analyses were employed. The five-factor model received statistical support from the CFA. Measurement invariance analyses showed that strict invariance held for gender. Kapoor et al. (2023) addressed the structural and generalizability aspects of validity using data obtained from 15,868 respondents. Note that the data were collected from university students in the United States in previous studies. The CFA provided supporting evidence for the five-factor model. Supporting evidence for strict invariance was received for ethnicity (i.e., European Americans, African American, Asian American, and Hispanic American). Finally, Brauer

et al. (2022) conducted a CFA using the responses of 511 German respondents and measurement invariance analyses using the responses of 502 German respondents. The findings showed that the K-DOCS was a five-factor instrument and that scalar invariance held for gender.

The external aspect of validity was also examined for K-DOCS scores. Kaufman (2012) estimated correlations among the five subscales and five personality traits (Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Openness to Experience) to examine convergent validity. Consistent with previous research (Feist, 2019a, 2019b; Grajzel et al., 2023; King et al., 1996), statistically significant ($p < .001$) positive correlation coefficients were estimated between extraversion and the Self/Everyday ($r = .19$), Scholarly ($r = .13$), and Performance ($r = .24$) subscales as well as between openness to experience and the Self/Everyday ($r = .15$), Scholarly ($r = .42$), Performance ($r = .31$), and Artistic ($r = .19$) subscales.

McKay et al. (2017) and Miroshnik et al. (2022) also examined the relationship between the five subscales of the K-DOCS and certain personality traits. McKay et al. (2017) conducted a study with 802 American and 500 Polish respondents. The correlation coefficients among the five K-DOCS subscales and extraversion as well as openness to experience were statistically significant at the $p < .05$ level, except for the coefficient between extraversion and the Scholarly subscale. The coefficients ranged from .12 to .51. Miroshnik et al. (2022) collected data from 1,313 Russian respondents to examine the relationship between scores on K-DOCS and scores on the short version of the Creative Behavior Inventory (Dollinger, 2003) as well as certain personality traits (e.g., curiosity). Statistically significant ($p < .001$) positive correlation coefficients were estimated between the following: the Artistic subscale and the Visual arts ($r = .54$) as well as Crafts ($r = .30$) subscales of the inventory; the Performance subscale and the Literature ($r = .43$) subscale of the inventory; and the Scholarly subscale and curiosity ($r = .24$).

Kandemir and Kaufman (2020) worked with 1,215 university students in Türkiye to compare K-DOCS across different majors. The students were pursuing degrees in various fields (e.g., Primary Education, Social Science Education, and Mathematics). After conducting two-way ANOVA analyses, Kandemir and Kaufman (2020) found no statistically significant differences among majors for the Self/Everyday and Scholarly subscales. On the other hand, in certain subscales, students majoring in art (the Performance and Artistic subscales) and mathematics (the Scientific subscale) scored significantly higher, providing evidence for convergent validity.

Validity of the 42-Item version

Şahin (2016a) conducted a study with two different samples. One sample consisted of 241 and the other sample consisted of 254 Turkish high school students. Şahin ran an EFA and a CFA. After conducting the EFA on the responses from the first sample, Items 1, 5, 16, 28, 33, 41, 42, and 49 were removed from the scale due to low factor loadings or loading on more than one factor. The CFA conducted on the responses from the second sample provided supporting evidence for the five-factor and 42-item scale.

Validity of the 20-Item version

Tan et al. (2021) worked with three different undergraduate student samples in Malaysia. Initially, an EFA was conducted using 484 responses for the 50-item version. Based on the results, four items with the highest factor loadings were retained for each subscale, resulting in 20 items in total. Afterwards, the 20-item and the 50-item versions were compared based on model fit indices. Responses obtained from 724 students were analyzed through a CFA. The 20-item version had superior model fit indices compared to the 50-item version. Measurement invariance was also tested for ethnicity (Malay vs. non-Malay) in the second sample. Supporting evidence for strict invariance was obtained. Finally, Tan et al. collected data from 201 students using the 20-item version and examined convergent validity. Correlation coefficients between scores on the 20-item version and scores on the Creative Self-Efficacy (Beghetto, 2006), Self-Perceived Creativity (Zhou & George, 2001), story writing, and Product Improvement Task (Torrance, 1974) were estimated. Statistically significant ($p < .001$) positive correlation coefficients were obtained between scores on the K-DOCS and scores on the Creative Self-Efficacy (r s ranged from .19–.39) as well as Self-Perceived Creativity (r s ranged from .26–.41).

Reliability of the original version

A range of reliability coefficients were reported for K-DOCS scores. Kaufman (2012) estimated Cronbach's alpha (α ; Cronbach, 1951) coefficients for internal consistency and test–retest reliability coefficients for temporal stability. The α coefficients were as follows: .83 (Artistic), .86 (Self/Everyday, Scholarly, and

Scientific), and .87 (Performance). The test–retest reliability coefficients were .80 (Self/Everyday), .76 (Scholarly), .86 (Performance), .78 (Scientific), and .81 (Artistic). Kaufman (2012) did not report the reliability coefficients for the total score due to the multidimensional structure of the scale (J. C. Kaufman, personal communication, July 27, 2022). The reliability coefficients reported in other studies for the full version were between .69 and .98 for the total score and between .63 and .94 for the subscale scores.

Reliability of the short versions

The reliability of scores on the short versions of the K-DOCS was also acceptable. The α coefficients for K-DOCS total and subscale scores were between .77 and .90 in Şahin (2016a). The α coefficients reported in other studies ranged from .76–.92 for the 42-item version. The α coefficients for scores on the subscales were between .67 and .86 in the study by Tan et al. (2021). The α coefficients reported in other studies were between .64 and .86 for the 20-item version.

Variation among the reliability coefficients

As was just presented, there is a considerable variation in the reliability of K-DOCS scores. This variation can be attributed to two distinct sources of measurement error: random error and systematic error (American Educational Research Association et al., 2014). Random error, stemming from unpredictable factors (e.g., health issues of respondents), contributes to the scattered distribution of scores around the true score (American Educational Research Association et al., 2014). This type of error could lead to inconsistent results upon repeated administration of an instrument. On the other hand, systematic error, which arises due to consistent biases in measurement, can be influenced by various factors such as culture, ethnicity, or gender (American Educational Research Association et al., 2014).

A quick investigation of studies on the K-DOCS reveals that the studies differed from each other with respect to certain variables, including Caucasian percentage, continent, country, culture, educational level, ethnicity, female percentage, language, mean of the age of respondents, mean of the total score, population, publication type (i.e., published vs. unpublished), publication year, scale version (i.e., full vs. short), *SD* of the age of respondents, and study focus (i.e., applied vs. psychometric). There is a possibility that some of these variables significantly impact the reliability of K-DOCS scores and explain a certain amount of systematic error because the reliability of scores on an instrument may vary across subgroups (e.g., different countries and ethnicities) and may be affected by certain variables (e.g., Deng et al., 2019; Graham et al., 2006, 2011; Vicent et al., 2019).

It should be noted that several studies did not estimate reliability coefficients for the K-DOCS and committed reliability induction (i.e., the practice of not reporting reliability coefficients). There are two types of induction: reliability induction by report (i.e., reporting the reliability estimates of prior research) and reliability induction by omission (i.e., reporting no reliability estimate). Note that both types of induction are discouraged, and they warrant caution among researchers. That is because reliability induction has the potential to lead to erroneous interpretations of score reliability for a given instrument and a large number of errors in reliability generalization meta-analysis studies (Sánchez-Meca et al., 2021).

PURPOSE OF THE STUDY

The K-DOCS has been examined in different aspects, such as its factorial structure (e.g., Şahin, 2016a), validity (e.g., Awofala & Fatade, 2015), and measurement invariance (e.g., Kapoor et al., 2023). However, no study has focused on the overall reliability coefficients and addressed the variability of the reliability of K-DOCS scores. This situation represents an important research gap because previous studies showed that the reliability of scores on an instrument varies across subgroups and is impacted by certain variables (Deng et al., 2019; Graham et al., 2006, 2011; Vicent et al., 2019). Moreover, there has been a lack of research exploring reliability induction for the K-DOCS and comparing the characteristics of studies that omitted reliability with those that reported reliability. Due to the absence of evidence, it is not certain whether the inducing and reporting studies have similar characteristics or whether there is a reporting bias problem with regard to the reliability of K-DOCS scores (Sánchez-Meca et al., 2021).

In the present study, we aim to address these issues using the reliability generalization approach (Henson & Thompson, 2002; Vacha-Haase, 1998). Reliability generalization (RG) is a method that allows researchers to estimate an overall reliability coefficient, examine reliability for different subgroups, identify the sources of the variability in the reliability coefficients, and investigate reliability induction for an instrument (Sánchez-Meca et al., 2021; Vacha-Haase, 1998).

The following research questions are answered in the present study:

1. What are the pooled α coefficients for the K-DOCS total score and subscale scores?
2. How do the characteristics of the samples (Caucasian percentage, continent, country, culture, educational level, ethnicity, female percentage, language, mean of the age, *SD* of the age, mean of the total score, and population), research (publication type, publication year, and study focus), and instrument (scale version) impact the overall reliability coefficients?
3. Are the characteristics of the inducing studies different from the characteristics of the reporting studies with respect to the mean of the total score, means and *SDs* of the sample age, and percentages of female and Caucasian?

METHOD

SEARCH STRATEGIES

In order to find the empirical studies on the K-DOCS, we conducted an internet search in several databases including Academic Search Complete, ERIC, ProQuest, PsycARTICLES, PsycINFO, and Web of Science. The following keywords were searched in the entire paper: “Kaufman Domains of Creativity Scale”, “K-DOCS and Kaufman,” and “K-DOCS.” Additionally, we conducted a search on Google Scholar using these keywords as well as the *cited by* option. Overall, 1001 studies were found.

In order to find additional studies, we scanned four prestigious journals on creativity. Those journals were *Creativity Research Journal*, *The Journal of Creative Behavior*, *Psychology of Aesthetics, Creativity, and the Arts*, and *Thinking Skills and Creativity*. We obtained an additional 26 studies through this process.

Finally, we conducted a backward search using the references cited in the reporting and inducing studies. We found four studies: Two of them were not written in English, while the other two were work-in-progress papers. As of September 1, 2022, the internet search produced 1031 studies that mentioned the K-DOCS. See Figure 1 for the Internet search procedure.

SELECTION CRITERIA

Studies were expected to meet certain criteria regarding research format, test version, language, publication year, and reliability coefficient to be included in the analyses. We included the empirical studies in which the 50-item version, the 42-item version, or the 20-item version were administered in the analyses. The reason is that these three versions were validated in previous studies. In some studies, researchers used the 50-item version but focused on certain subscales; we included those studies in the analyses as well because we conducted separate analyses for each subscale (see Table S1).

We selected the studies reported in English. The reasons are that most of the studies were published in English and that the authorship team does not speak any other languages besides their mother tongue. We kept the studies that have been conducted since 2012 because Kaufman (2012) published an article on the development of the K-DOCS in that year. See Table S1 for the publication years of the studies.

Finally, we kept the studies that reported α coefficients. Initially, we also intended to conduct RG analyses for other types of reliability coefficients. However, the overwhelming majority of the studies (87.5%) reported α coefficients. Only seven studies reported other types of reliability coefficients, including test–retest reliability coefficients (four studies) and omega coefficients (three studies). Separate RG analyses could not be conducted for test–retest and omega coefficients, and thus, those seven studies were excluded.

SELECTION OF STUDIES

Duplicates, nonempirical studies, and non-English studies were excluded. We kept 471 studies for consideration. Three hundred and sixty-five studies only mentioned the K-DOCS, and thus, they were excluded. The scale was administered in 106 studies. However, the reliability coefficients were not estimated in all 106 studies. Upon an initial examination of the studies, we identified 62 studies in which a reliability coefficient was reported.

The full version (original or translated) was used in 41 studies, the 42-item version was used in six studies, and the 20-item version was used in two studies. In nine studies, either one subscale or a few subscales were administered. In one study, respondents were given 54 items. Finally, a researcher-modified version of the scale was used in five studies.

We sent emails to the corresponding authors of the papers that did not include reliability coefficients for the entire scale and/or for each subscale. Fourteen corresponding authors emailed back and shared their

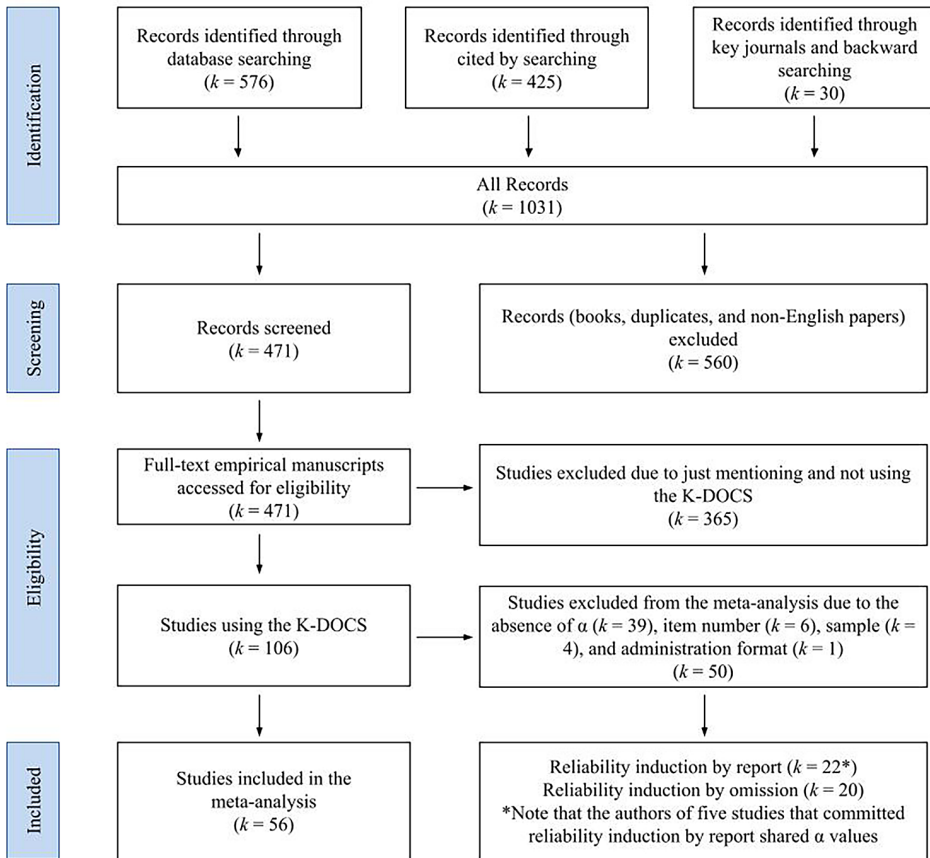


FIGURE 1. PRISMA Flowchart.

estimations for either the K-DOCS total score or subscale scores—note that five of those authors committed reliability induction. We obtained an additional 39 reliability coefficients through email. Overall, there were 67 studies whose α coefficients were available.

The α coefficients of four studies (Awofala & Fatade, 2015; Lee & Portillo, 2022; Tan et al., 2016, 2021) whose samples were the same as another study were excluded—note that more than one set of α coefficients was reported in some studies. We did this to avoid the dependency problem because when multiple effect sizes are reported in the same study based on the same sample, a dependency problem may arise (van den Noortgate et al., 2015). The α values were calculated from different samples in the rest of the studies that reported multiple α coefficients. Thus, we treated those studies as independent studies.

Five studies (Ashraf et al., 2019; Dong et al., 2022; Dousay & Weible, 2019; Jung et al., 2021; Magnusson, 2018) were removed from the analyses due to item number—note that the researchers used neither the full version nor the short versions and did not focus on any subscales in these studies. Susanto et al. (2018) administered 54 items, and this study was excluded. Another study (Aznar et al., 2021) was excluded because the parents filled out the K-DOCS for their children.

The final sample consisted of 56 studies. It should be noted that because some studies reported multiple α coefficients from independent samples, a total of 60 α values were obtained in this study. In the labeling of studies that reported more than one α coefficient, the source name was kept the same, with numerical identifiers added to the end (e.g., McKay et al., 2017_1 and McKay et al., 2017_2). It should be noted that the inducing studies were also saved in the coding sheet because we compared the inducing studies with the

reporting ones. Materials and analysis codes for the present study can be obtained by emailing the corresponding author.

ANALYSES

Publication bias

Publication bias was examined before the main analyses. We made use of Egger's regression test (Egger et al., 1997) as well as Begg and Mazumdar's (1994) rank correlation test— p -values greater than .05 indicate a lack of publication bias. We also utilized the trim-and-fill method (Duval & Tweedie, 2000) and the funnel plot approach.

Reliability generalization meta-analyses

Before conducting the RG analyses, the α coefficients were converted to Bonett's T -values (Bonett, 2002b). This procedure was employed to normalize the distributions of the α coefficients. We used the following formula for the transformation: $L_i = \ln(1-|\alpha_i|)$.

We conducted random-effects meta-analyses with inverse variance weights (Vacha-Haase, 1998) to estimate the overall α values for K-DOCS total scores and subscale scores. Residual heterogeneity among the α coefficients was examined with Q -statistics (Cochran, 1954) and I^2 values (Higgins et al., 2003). High I^2 values and a statistically significant Q -test indicate heterogeneity. As suggested by Faggion Jr. et al. (2021), prediction intervals (PI) were also estimated for the pooled reliability estimates under the random-effects model. PIs show the range of the predicted reliability coefficients in a future study and are estimated based on the population of studies that are already included in the meta-analysis (Nagashima et al., 2019).

We conducted moderator analyses to examine the effect of several variables on the α coefficients. The studies included in the moderator analyses were coded based on the coding scheme seen in Table S2. There were 10 categorical moderator variables: continent, country, culture, educational level, ethnicity, language, publication type, population, scale version, and study focus. The subcategories of the categorical variables can be seen in Table S2. There were six continuous moderator variables: publication year, female percentage, Caucasian percentage, mean of the age, SD of the age, and mean of the total score (see Tables S1 and S3 for details). We performed moderator analyses for both the categorical and continuous variables using univariate meta-regression models as implemented in the *metafor* package (Viechtbauer, 2010). The categorical variables were recoded into dichotomous dummy variables for conducting categorical moderator analyses.

Some studies failed to report the values of certain moderator variables. No action is taken for those missing values, and the missing values were coded as missing. All the analyses were conducted in R with the *metafor* package.

Reliability induction

In order to evaluate reliability induction by report and reliability induction by omission, we divided the studies into categories based on estimating, reporting, and omitting reliability coefficients. Then, we calculated the reliability induction rates for the K-DOCS. Afterward, the inducing studies and reporting studies were compared with an independent-samples t -test with respect to the total score, means and SD s of the sample age, and percentage of females and Caucasians.

RESULTS

CHARACTERISTICS OF THE STUDIES

Tables S1 through S4 provide information about the characteristics of the studies ($k = 56$). Table S1 shows the publication year (between 2012 and 2022), number of items (between 9 and 50), number of respondents (between 34 and 2,318), and α coefficients for the total score (between .69 and .99) as well as the subscale scores (between .64 and .95) for each study. Table S2 presents the following categorical study characteristics: test version (i.e., original, validated translation, validated short, and mere translation) used in each study, as well as the publication type (i.e., article, thesis, or conference paper), location (i.e., country and continent), population, language, educational level of the sample, ethnicity percentages of the sample, culture (i.e., East or West), and focus (i.e., applied or psychometric) for each study.

The non-U.S. countries include Türkiye (15.00%), China (6.67%), Malaysia (6.67%), Nigeria (5.00%), Kazakhstan (3.33%), Australia (1.67%), Croatia (1.67%), Czech Republic (1.67%), Lithuania (1.67%), Mexico (1.67%), Myanmar (1.67%), Pakistan (1.67%), Poland (1.67%), Romania (1.67%), South Korea (1.67%), Taiwan (1.67%), and the Netherlands (1.67%). Note that 3.33% of the studies were conducted in

more than one country. The non-English versions of the K-DOCS were administered in Turkish (15%), Chinese (8.33%), Kazakh (3.33%), Nigerian (3.33%), Burmese (1.66%), Croatian (1.67%), Czech (1.67%), Dutch (1.67%), Korean (1.67%), Lithuanian (1.67%), Malay (1.67%), Polish (1.67%), Romanian (1.67%), Spanish (1.67%), and Urdu (1.67%). The Eastern countries include China, Kazakhstan, Malaysia, Myanmar, Nigeria, Pakistan, South Korea, Taiwan, and Türkiye; while the Western countries include Australia, Austria, Canada, Croatia, Czech Republic, Lithuania, Mexico, Poland, Romania, the Netherlands, and the United States.

Table S3 provides an overview of several continuous variables. The number of respondents ranged from 34 to 2,318 ($M = 412.18$, $SD = 387.34$). In total, there were 24,731 respondents. The mean age was between 17.85 and 57.59 ($M = 24.99$, $SD = 7.68$). The SD of the mean age ranged from 0.36 to 16.20 ($M = 5.31$, $SD = 4.34$). Female percentages ranged from 17 to 97.60 ($M = 62.03$, $SD = 16.42$). The percentages of Caucasians were between 0 and 100 ($M = 48.93$, $SD = 42.34$). Finally, the number of items used in 60 K-DOCS administrations ranged from 9 to 50 ($M = 42.63$, $SD = 14.09$).

PUBLICATION BIAS

The trim-and-fill procedure implied that the Self/Everyday subscale and the Scientific subscale should have nine and four α coefficients imputed to them, respectively. No imputation was needed for the total scale and the other three subscales. On the other hand, the funnel plots indicated no publication bias. The funnel plots are presented in the supplementary material (see Figures S1–S6).

Tests based on rank correlation and regression analyses also indicated a lack of publication bias. The rank correlation analyses provided nonsignificant Kendall's τ values between the α coefficients and the corresponding standard errors of the total scale as well as the subscales. The Kendall's τ values were not statistically significant for the total scale ($\tau = -0.220$, $p = .101$) as well as for the Self/Everyday ($\tau = -0.195$, $p = .059$), Scholarly ($\tau = -0.139$, $p = .182$), Performance ($\tau = -0.011$, $p = .919$), Scientific ($\tau = -0.139$, $p = .193$), and Artistic ($\tau = -0.048$, $p = .657$) subscales. The Egger's tests of symmetry for the four funnel plots were not statistically significant for the total scale ($z = -1.208$, $p = .238$) as well as for the Scholarly ($z = -0.509$, $p = .613$), Performance ($z = 0.059$, $p = .953$), and Artistic ($z = 0.723$, $p = .474$) subscales. However, statistically significant results were obtained for the Self/Everyday ($z = -2.202$, $p = .033$) and Scientific ($z = -2.163$, $p = .031$) subscales based on Egger's tests of symmetry. In sum, we concluded that publication bias did not pose a substantial risk to the reliability of K-DOCS scores.

MEAN AND POOLED RELIABILITY COEFFICIENTS

As shown in Table 1, 28 α coefficients were available for K-DOCS total scores. It should be noted that not all studies reported reliability coefficients for the total scale, as some studies reported α coefficients for only the subscales. The numbers of α coefficients obtained for the subscales were 45 (Self/Everyday), 44 (Scholarly), 41 (Performance), 42 (Scientific), and 42 (Artistic). First, we calculated the mean α value based on the α values reported in the studies (i.e., raw α coefficients) without applying any transformation. The mean α value of the raw coefficients was .887 ($SD = 0.066$) for the K-DOCS total score. The mean α values

TABLE 1. Mean α coefficients and heterogeneity statistics

	k	α_+	95% CI, LL; UL	95% PI, LL; UL	Q	I^2	τ^2
Total	28	.904	.880; .923	.686; .971	1369.628***	98.50	.353
Self/Everyday	45	.825	.805; .844	.636; .916	1124.650***	96.80	.136
Scholarly	44	.858	.841; .873	.707; .931	1140.890***	96.84	.134
Performance	41	.887	.877; .896	.811; .932	541.145***	93.98	.067
Mechanical/Scientific	42	.867	.855; .879	.768; .924	625.307***	94.77	.079
Artistic	42	.861	.848; .873	.756; .921	530.694***	94.85	.080

Note. k is the number of α values; α_+ is the mean coefficient α ; CI is confidence interval; PI is prediction interval; LL is lower limit of 95% confidence interval for α_+ ; UL is upper limit of 95% confidence interval for α_+ ; Q is heterogeneity statistic, $df = (n - 1)$; τ^2 is estimated total heterogeneity; I^2 is heterogeneity index. *** $p < .001$.

AN RG META-ANALYSIS OF THE K-DOCS

calculated for the Self/Everyday (.812, $SD = 0.066$), Scholarly (.846, $SD = 0.067$), Performance (.882, $SD = 0.029$), Scientific (.860, $SD = 0.044$), and Artistic (.854, $SD = 0.049$) subscales were also over .800.

Second, transformed α values based on Bonett (2002a) were used with random-effects meta-analysis to estimate pooled α coefficients (from this point on, the term pooled α is used). Table 1 presents the pooled α coefficients estimated for the K-DOCS total score as well as the subscale scores. Under the random-effects model, the pooled α value for the K-DOCS total score was estimated to be .904 (95% CI: [.880, .923] and PI: [.686, .971]). The pooled α value was statistically significant for the total score at the $p < .001$ level. The pooled α values for Self/Everyday (.825, 95% CI: [.805, .844] and PI: [.636, .916]); Scholarly (.858, 95% CI: [.841, .873] and PI: [.707, .931]); Performance (.887, 95% CI: [.877, .896] and PI: [.811, .932]); Scientific (.867, 95% CI: [.855, .879] and PI: [.768, .924]); and Artistic (.861, 95% CI: [.848, .873] and PI: [.756, .921]) subscale scores were also statistically significant at the $p < .001$ level.

MODERATOR VARIABLES

The Q -test statistics (see Table 1) and I^2 values ($> 90\%$) denoted considerable heterogeneity among the α coefficients for the K-DOCS total score as well as the subscale scores. The variability detected among the α coefficients can also be seen in the forest plots (see Figures S7–S12). To examine potential sources of this heterogeneity, we carried out moderator analyses. The results of the moderator analyses are displayed in Tables 2–8.

TABLE 2. Results of the categorical moderator analyses for the total scale

Variable	Category	k	α	95% LL	95% UL	Q_B	p
Publication Type	Published	24	.899	.872	.920	1.474	.225
	Unpublished	4	.932	.877	.962		
Continent	Africa	3	.829	.691	.906	6.616	.085
	Asia	4	.914	.856	.949		
	Europe	8	.897	.852	.928		
	N. America	12	.926	.900	.945		
Country	U.S.	11	.923	.891	.946	2.552	.110
	non-U.S.	17	.890	.854	.916		
Culture	East	13	.896	.859	.924	1.174	.278
	West	14	.918	.890	.939		
Ethnicity	African	3	.830	.666	.913	3.385	.336
	Asian	6	.916	.865	.948		
	Caucasian	8	.897	.845	.932		
	Mixed	10	.912	.873	.939		
Language	English	13	.910	.875	.936	0.280	.596
	non-English	15	.898	.862	.925		
Population	Adults	3	.878	.753	.939	1.108	.575
	Students	20	.902	.872	.925		
	General	5	.923	.868	.955		
Education	Undergraduate	12	.908	.878	.930	0.591	.744
	High school	4	.889	.816	.932		
	Mixed	7	.894	.847	.923		
Test version	Original	12	.913	.876	.939	0.743	.863
	Valid Trans.	2	.888	.732	.953		
	Valid Short1	4	.908	.830	.950		
	Mere Trans.	10	.893	.842	.928		
Focus	Applied	24	.902	.874	.923	0.294	.588
	Psychometrics	4	.918	.851	.955		

Note. k is the number of α values; LL is the lower limit of 95% confidence interval; UL is the upper limit of 95% confidence interval; Q_B is the heterogeneity statistic.

TABLE 3. Results of the categorical moderator analyses for the Self/Everyday Subscale

Variable	Category	<i>k</i>	α	95% LL	95% UL	Q_B	<i>p</i>
Publication Type	Published	37	.827	.804	.847	0.162	.687
	Unpublished	8	.816	.759	.860		
Continent	Asia	9	.833	.785	.870	2.961	.228
	Europe	12	.793	.742	.833		
	N. America	20	.836	.805	.861		
Country	U.S.	19	.834	.803	.860	0.766	.382
	non-U.S.	25	.816	.786	.842		
Culture	East	16	.818	.780	.850	0.154	.695
	West	27	.827	.799	.851		
Ethnicity	Asian	5	.853	.795	.894	1.653	.434
	Caucasian	13	.809	.766	.845		
	Mixed	23	.824	.795	.849		
Language	English	24	.823	.793	.848	0.022	.882
	non-English	20	.826	.793	.853		
Population	Adults	5	.851	.791	.894	1.138	.566
	Students	35	.820	.796	.841		
	General	5	.833	.766	.881		
Education	Undergraduate	22	.818	.787	.845	3.078	.215
	High school	6	.791	.717	.845		
	Mixed	10	.849	.809	.880		
Test version	Original	20	.827	.796	.853	6.915	.141
	Valid Trans.	7	.831	.777	.871		
	Valid Short1	4	.793	.702	.856		
	Valid Short2	3	.719	.573	.815		
	Mere Trans.	9	.846	.804	.879		
Focus	Applied	35	.828	.805	.849	0.277	.599
	Psychometrics	10	.815	.766	.854		

Note. *k* is the number of α values; LL is the lower limit of 95% confidence interval; UL is the upper limit of 95% confidence interval; Q_B is the heterogeneity statistic.

Scholarly subscale

For the Scholarly subscale, only educational level was a statistically significant moderator variable ($Q_B = 9.898$, $df = 3$, $p < .05$). The pooled α value estimated for the samples with graduate students (.893) was higher than the pooled α value calculated for the mixed samples (.890). The pooled α values estimated for these two subgroups were higher than the α values estimated for the other subgroups (undergraduate [.852] and high school [.812]). Table 4 shows the full results for the Scholarly subscale.

Categorical variables

Regarding the categorical variables, none of the moderator variables significantly impacted the α coefficients for K-DOCS total scores and the Self/Everyday subscale scores. In other words, the α coefficients estimated for each subgroup did not differ from one to another for K-DOCS total scores and Self/Everyday subscale scores. The findings can be seen in Tables 2 and 3.

Performance subscale

Educational level was also a statistically significant moderator variable for the Performance subscale ($Q_B = 9.924$, $df = 2$, $p < .01$). The pooled α value for the samples composed of mixed educational levels (.903) was higher than the pooled α values for the samples composed of undergraduate (.883) or high school levels (.857). Continent ($Q_B = 12.802$, $df = 2$, $p < .01$) and ethnicity ($Q_B = 12.811$, $df = 2$, $p < .01$) were the other statistically significant moderators. The pooled α value for the studies conducted in North America (.896) was higher than the α values for the studies conducted in Asia (.895) or Europe (.858). The

TABLE 4. Results of the categorical moderator analyses for the Scholarly Subscale

Variable	Category	<i>k</i>	α	95% LL	95% UL	Q_B	<i>p</i>
Publication Type	Published	37	.857	.839	.874	0.036	.850
	Unpublished	7	.862	.816	.896		
Continent	Asia	9	.848	.809	.879	1.586	.452
	Europe	12	.850	.816	.877		
	N. America	19	.869	.846	.888		
Country	U.S.	18	.867	.842	.888	1.012	.314
	non-U.S.	25	.851	.826	.871		
Culture	East	16	.853	.823	.878	0.072	.789
	West	26	.858	.835	.877		
Ethnicity	Asian	5	.872	.823	.908	0.578	.749
	Caucasian	13	.852	.818	.879		
	Mixed	22	.859	.835	.880		
Language	English	23	.854	.829	.875	0.201	.654
	non-English	20	.862	.836	.883		
Population	Adults	4	.881	.826	.918	0.997	.608
	Students	35	.855	.835	.872		
	General	5	.863	.808	.902		
Education	Undergraduate	22	.852	.829	.872	9.898	.019*
	Graduate	2	.893	.827	.934		
	High school	6	.812	.751	.858		
	Mixed	8	.890	.860	.913		
Focus	Applied	34	.861	.809	.880	0.346	.556
	Psychometrics	10	.849	.842	.877		

Note. *k* is the number of α values; LL is the lower limit of 95% confidence interval; UL is the upper limit of 95% confidence interval; Q_B is the heterogeneity statistic. *Shows the significant *p* Values.

pooled α value for the Asian samples (.908) was higher than the α values for the ethnically mixed (.894) and Caucasian (.861) samples. The full results for the Performance subscale are presented in Table 5.

Scientific subscale

Ethnicity was also a statistically significant moderator variable for the Scientific subscale ($Q_B = 8.830$, $df = 2$, $p < .01$). The pooled α value for the ethnically mixed samples (.880) was higher than the α values for the Asian (.859) and Caucasian (.840) samples. The other statistically significant moderators were continent ($Q_B = 13.409$, $df = 2$, $p < .001$), country ($Q_B = 8.663$, $df = 1$, $p < .01$), language ($Q_B = 6.356$, $df = 1$, $p < .05$), and test version ($Q_B = 9.661$, $df = 1$, $p < .05$). The pooled α value for the North American samples (.885) was higher than the α values for the Asian (.854) and European (.838) samples. The pooled α value for the U.S. samples (.885) was higher than the α values for the non-U.S. samples (.851). Similarly, the pooled α value for the studies conducted with the English version (.879) was higher than the α value for the studies conducted with the non-English versions (.849). Finally, the pooled α value for the studies conducted with the original version (.881) was higher than the α values for the studies conducted with the other versions (valid translation, .859; mere translation, .855; valid short_1, .850; and valid short_2, .813). The full results for the Scientific subscale are presented in Table 6.

Artistic subscale

For the Artistic subscale, statistical significance was detected for continent ($Q_B = 14.235$, $df = 2$, $p < .001$), country ($Q_B = 5.826$, $df = 1$, $p < .05$), and language ($Q_B = 6.579$, $df = 1$, $p < .05$). The pooled α value for the North American samples (.877) was higher than the α values for the Asian (.867) and European (.822) samples. The pooled α value for the U.S. samples (.877) was higher than the α value for the non-U.S. samples (.847). The pooled α value for the English version (.874) was higher than the α value for the non-English versions (.842). The full results for the Artistic subscale are in Table 7.

TABLE 5. Results of the categorical moderator analyses for the Performance Subscale

Variable	Category	<i>k</i>	α	95% LL	95% UL	Q_B	<i>p</i>
Publication Type	Published	34	.885	.874	.895	0.805	.370
	Unpublished	7	.896	.873	.916		
Continent	Asia	7	.895	.875	.912	12.802	.002*
	Europe	11	.858	.836	.877		
	N. America	19	.896	.885	.907		
Country	U.S.	18	.895	.881	.908	2.583	.108
	non-U.S.	22	.880	.865	.893		
Culture	East	13	.881	.862	.897	0.406	.524
	West	26	.888	.875	.899		
Ethnicity	Asian	4	.908	.883	.927	12.811	.002*
	Caucasian	11	.861	.839	.879		
	Mixed	22	.894	.883	.904		
Language	English	23	.893	.881	.904	2.560	.110
	non-English	17	.877	.860	.892		
Population	Adults	4	.902	.872	.925	2.783	.249
	Students	33	.883	.872	.893		
	General	4	.902	.873	.925		
Education	Undergraduate	21	.883	.872	.894	9.924	.007*
	High school	6	.857	.828	.881		
	Mixed	8	.903	.886	.912		
Test version	Original	19	.894	.881	.906	8.856	.065
	Valid Trans.	5	.882	.852	.906		
	Valid Short1	4	.858	.817	.890		
	Valid Short2	3	.849	.797	.887		
	Mere Trans.	8	.894	.873	.912		
Focus	Applied	31	.888	.877	.899	0.191	.662
	Psychometrics	10	.883	.862	.901		

Note. *k* is the number of α values; LL is the lower limit of 95% confidence interval; UL is the upper limit of 95% confidence interval; Q_B is the heterogeneity statistic. *Shows the significant *p* Values.

Continuous variables

With respect to the continuous variables, statistical significance was not detected for any of the variables for K-DOCS total scores, Self/Everyday subscale scores, and Scientific subscale scores (see Table 8). On the other hand, the mean age of the study sample was found to be a statistically significant predictor of the α values for the Scholarly ($p = .019$), Performance ($p = .001$), Scientific ($p = .019$), and Artistic ($p = .046$) subscales. *SD* of the sample age was a statistically significant moderator for the Scholarly ($p = .038$), Scientific ($p = .010$), and Performance ($p = .001$) subscales. Finally, the percentage of Caucasians and total mean were statistically significant moderators for the Performance ($p = .024$) and Artistic ($p = .017$) subscales, respectively.

RELIABILITY INDUCTION

Reliability induction was committed in several studies in which the K-DOCS was administered. Twenty-two studies reported the reliability coefficients estimated in previous studies and committed reliability induction by report. Most of these studies reported the α coefficients estimated by Kaufman (2012). On the other hand, 20 studies did not mention the reliability of K-DOCS scores and committed reliability induction by omission. The induction rates were 39.62% (total induction rate), 20.75% (induction by report), and 18.87% (induction by omission).

We examined the characteristics of the inducing studies, as suggested by Sánchez-Meca et al. (2021). The following variables were investigated for the inducing studies: the publication type, country, continent, culture, year of publication, focus of the study, language, scale version, population type of the sample,

TABLE 6. Results of the categorical moderator analyses for the Mechanical/Scientific Subscale

Variable	Category	<i>k</i>	α	95% LL	95% UL	Q_B	<i>p</i>
Publication Type	Published	35	.862	.849	.875	3.410	.065
	Unpublished	7	.890	.863	.911		
Continent	Asia	8	.854	.826	.878	13.409	.001*
	Europe	11	.838	.811	.862		
	N. America	19	.885	.871	.898		
Country	U.S.	18	.885	.869	.898	8.663	.003*
	non-U.S.	23	.851	.834	.867		
Culture	East	14	.850	.825	.870	3.381	.066
	West	26	.874	.860	.887		
Ethnicity	Asian	5	.859	.821	.888	8.830	.012*
	Caucasian	11	.840	.813	.864		
	Mixed	22	.880	.866	.893		
Language	English	23	.879	.865	.892	6.356	.012*
	non-English	18	.849	.828	.867		
Population	Adults	4	.875	.833	.907	0.495	.781
	Students	34	.865	.851	.878		
	General	4	.876	.833	.908		
Education	Undergraduate	21	.863	.845	.879	3.826	.148
	High school	6	.843	.802	.876		
	Mixed	9	.883	.858	.903		
Test version	Original	19	.881	.866	.895	9.661	.047*
	Valid Trans.	5	.859	.821	.889		
	Valid Short1	4	.850	.804	.885		
	Valid Short2	3	.813	.746	.863		
Focus	Mere Trans.	9	.855	.827	.879	0.056	.813
	Applied	31	.868	.853	.881		
	Psychometrics	11	.865	.839	.886		

Note. *k* is the number of α values; LL is the lower limit of 95% confidence interval; UL is the upper limit of 95% confidence interval; Q_B is the heterogeneity statistic. *Shows the significant *p* Values.

ethnicity, mean and *SD* of the sample age, educational level, and the percentage of females and Caucasian. The inducing studies were published between 2014 and 2022. The studies consisted of journal articles (53.84%) and unpublished studies (46.16%), including theses, proceedings, abstracts, and technical reports. The studies were carried out in several countries, including the United States (23.10%), the United Kingdom (12.80%), China (5.10%), Czech Republic (5.10%), the Netherlands (5.10%), Argentina (2.10%), Canada (2.60%), Finland (2.10%), Germany (2.10%), India (2.10%), Latvia (2.10%), Malaysia (2.10%), Mexico (2.10%), Pakistan (2.10%), Singapore (2.10%), Sweden (2.10%), and Türkiye (2.10%). Approximately 5% of the studies were conducted in more than one country. The country could not be determined for five studies (12.80%).

The samples were composed of Asian (12.70%), Caucasian (20.60%), or ethnically mixed (28.20%) respondents. Approximately 39% of the studies did not specify the ethnicities of the respondents. Students (56.40%), individuals from the general population (34.90%), or a combination of these two groups (7.70%) comprised the respondents. With respect to education, undergraduate (43.60%), high school (2.60%), middle school (2.60%), or mixed levels of (25.60%; e.g., undergraduate and graduate students) students were the respondents. Both the original version (46.20%) and translated versions were used in the inducing studies (33.30%). Approximately 21% of the studies did not mention the language in which the scale was administered. The K-DOCS was administered in psychometric studies (7.7%) as well as applied studies (92.3%).

We used an independent-samples *t*-test to compare the inducing and reporting studies for certain continuous variables. Those variables included the mean of the total score, the means and *SD*s of the sample

TABLE 7. Results of the categorical moderator analyses for the Artistic Subscale

Variable	Category	<i>k</i>	α	95% LL	95% UL	Q_B	<i>p</i>
Publication Type	Published	35	.857	.842	.870	2.326	.127
	Unpublished	7	.881	.852	.904		
Continent	Asia	8	.867	.841	.889	14.235	.0008*
	Europe	11	.822	.791	.848		
	N. America	19	.877	.862	.891		
Country	U.S.	18	.877	.860	.892	5.826	.016*
	non-U.S.	23	.847	.828	.864		
Culture	East	14	.846	.821	.868	2.223	.136
	West	26	.867	.851	.881		
Ethnicity	Asian	4	.862	.822	.894	4.871	.088
	Caucasian	12	.844	.818	.857		
	Mixed	22	.873	.859	.887		
Language	English	23	.874	.859	.888	6.579	.010*
	non-English	18	.842	.820	.861		
Population	Adults	4	.879	.837	.909	1.619	.445
	Students	33	.857	.841	.870		
	General	5	.873	.835	.902		
Education	Undergraduate	21	.858	.842	.872	5.155	.076
	High school	6	.831	.793	.861		
	Mixed	8	.875	.852	.894		
Test version	Original	19	.874	.856	.889	8.752	.068
	Valid Trans.	5	.830	.782	.868		
	Valid Short1	4	.828	.773	.870		
	Valid Short2	3	.832	.768	.878		
	Mere Trans.	9	.871	.845	.893		
Focus	Applied	32	.864	.848	.878	0.893	.345
	Psychometrics	10	.850	.820	.875		

Note. *k* is the number of α values; LL is the lower limit of 95% confidence interval; UL is the upper limit of 95% confidence interval; Q_B is the heterogeneity statistic. *Shows the significant *p* Values.

age, and the percentages of females and Caucasians. As seen in Table S4, no statistically significant differences were found between the inducing and reporting studies for any of the variables.

DISCUSSION

Reliability signifies the consistency and dependability of the measurements an instrument produces. The impact of sample characteristics (e.g., culture, ethnicity, and gender) and instrument characteristics (e.g., full vs. short version) on reliability is inevitable (American Educational Research Association et al., 2014). Nevertheless, it is imperative to emphasize that reliability must remain a non-negotiable attribute of scores on any instrument, regardless of the systematic impact of certain variables in favor of certain subgroups. In the case of significant impacts of certain variables on the reliability of scores on an instrument, RG analyses should be conducted to identify potential sources of systematic error. In the present study, we conducted RG analyses and focused on the reliability of K-DOCS scores, investigated variables that impact the reliability coefficients in a significant way across samples, and examined reliability induction.

OVERALL RELIABILITY COEFFICIENTS FOR THE K-DOCS

The RG meta-analyses were employed to answer the first question, which addresses the overall α values. The mean α values estimated in the present study are lower than the raw α coefficients reported by Kaufman (2012). Nevertheless, the mean α values for K-DOCS total scores as well as subscale scores are still over .80 and denote good reliability (O'Rourke et al., 2005). The findings indicate that the error (i.e., random

TABLE 8. Results of the continuous moderator analyses for the total scale and subscales

Moderator Variable	<i>k</i>	<i>b_j</i>	SE	<i>p</i>	<i>R</i> ²	<i>Q_E</i>
Total Scale						
Year	28	0.055	0.042	.191	.026	1369.272***
Female %	26	-0.008	0.007	.255	.001	1355.244***
Caucasian %	25	-0.002	0.003	.499	.001	1037.577***
Mean of the age (in years)	16	-0.026	0.024	.282	.008	875.680***
SD of the age (in years)	12	-0.020	0.038	.598	.001	845.208***
Total mean	16	-0.141	0.385	.715	.001	385.772***
Self/Everyday Subscale						
Year	45	-0.001	0.022	.978	.001	1106.294***
Female %	44	-0.001	0.004	.896	.001	1106.528***
Caucasian %	40	-0.002	0.002	.221	.014	934.224***
Mean of the age (in years)	32	0.005	0.007	.527	.001	584.412***
SD of the age (in years)	28	0.015	0.014	.276	.003	490.915***
Total mean	27	0.158	0.143	.271	.015	397.055***
Scholarly Subscale						
Year	44	-0.006	0.022	.788	.001	1138.833***
Female %	42	0.001	0.004	.789	.001	1125.036***
Caucasian %	39	0.001	0.001	.849	.001	881.757***
Mean of the age (in years)	29	0.017	0.007	.019*	.144	610.725***
SD of the age (in years)	25	0.031	0.015	.038*	.124	536.763***
Total mean	27	0.097	0.330	.768	.001	542.512***
Performance Subscale						
Year	41	-0.001	0.017	.931	.001	534.118***
Female %	40	-0.001	0.003	.618	.001	523.794***
Caucasian %	36	-0.002	0.001	.024*	.118	418.220***
Mean of the age (in years)	28	0.019	0.005	.001*	.339	287.406***
SD of the age (in years)	24	0.036	0.010	.001*	.366	250.569***
Total mean	24	-0.077	0.231	.739	.001	265.933***
Mechanical/Scientific Subscale						
Year	42	-0.006	0.017	.729	.001	620.913***
Female %	41	0.001	0.003	.712	.001	618.608***
Caucasian %	37	-0.001	0.001	.255	.009	469.468***
Mean of the age (in years)	28	0.015	0.006	.019*	.163	391.437***
SD of the age (in years)	24	0.029	0.011	.010*	.209	324.896***
Total mean	25	0.142	0.181	.432	.001	341.189***
Artistic Subscale						
Year	42	0.003	0.018	.845	.001	490.214***
Female %	40	-0.001	0.003	.814	.001	491.447***
Caucasian %	37	-0.001	0.001	.725	.001	444.413***
Mean of the age (in years)	28	0.013	0.006	.046*	.120	342.220***
SD of the age (in years)	24	0.024	0.013	.056	.118	309.701***
Total mean	25	0.220	0.092	.017*	.203	215.758***

Note. *k* is the number of α coefficients; *b_j* is the unstandardized regression coefficient; SE is the standard error; *R*² is the proportion of variance explained; *Q_E* is the statistic to test for residual heterogeneity; *Shows the significant *p* Values. ****p* < .001.

and systematic errors) in K-DOCS scores is relatively low and that the scale provides adequately reliable measures of respondents' creative behaviors. Our findings suggest that the K-DOCS can be used for research purposes, as the mean α coefficients are over .80 (Nunnally & Bernstein, 1994).

IMPACT OF THE MODERATOR VARIABLES

Some variables may significantly affect the reliability of K-DOCS scores across samples and explain a certain amount of systematic error in scale scores. Ten categorical and six continuous variables were analyzed to investigate the sources of variance in the α coefficients and to answer the second research question, which addresses the effect of moderator variables on the mean α coefficients.

Prior to conducting moderator analyses, we focused on the variation in the reliability of K-DOCS scores. The fluctuations in the α coefficients and CIs implied that the variation in the α values might be too high. To examine this issue, we ran heterogeneity analyses. The Q -test statistics and I^2 values estimated in the present study indicate a considerable amount of variation. This finding suggests that some moderator variables significantly impact the reliability of K-DOCS scores (Higgins et al., 2003; Thompson & Vacha-Haase, 2000). Thus, it is not sufficient and not advisable to report the reliability estimations of previous studies. Researchers using the scale should estimate reliability coefficients for each application because, as indicated by the PIs, the reliability coefficients for the total scale and subscales may be between .636 and .971 in future assessments.

Variables with a nonsignificant impact

None of the moderator variables affected the α coefficients for K-DOCS total scores and Self/Everyday subscale scores. Our findings imply that the reliability of total scores and Self/Everyday subscale scores does not significantly change across samples in terms of the moderator variables considered in the present study (Botella et al., 2010). This means that the moderator variables do not contribute to any systematic error in K-DOCS total scores and Self/Everyday subscale scores.

Unlike the other subscales, the Self/Everyday subscale does not require any training on a certain domain. Consider the following items: "Finding something fun to do when I have no money" and "Choosing the best solution to a problem" (Kaufman, 2012, p. 307). This situation negates the impact of formal and informal training on the Self/Everyday subscale items to a certain extent. The items on this subscale are possibly not as vulnerable as the items on the other subscales to the impact of training originated from certain variables, such as country, ethnicity, and educational level. This point is supported by a study (Kandemir & Kaufman, 2020) in which there were no differences among students from different majors regarding Self/Everyday subscale scores. This aspect of the Self/Everyday subscale may be a reason why the reliability coefficients for this subscale are not affected by the moderator variables. Note that the impact of training may have impacted the reliability of scores on other subscales to a certain extent.

Another reason for obtaining statistically nonsignificant results for the Self/Everyday subscale may be due to publication bias. Note that a statistically significant result was obtained for the Self/Everyday subscale based on Egger's tests of symmetry. Moreover, we found that the Self/Everyday subscale should have nine α coefficients imputed to it.

However, the Q -test statistics (see Table 4) and I^2 values ($> 90\%$) imply that some other variables may impact the reliability of total scores and Self/Everyday subscale scores. Specifically, the instructions given at the beginning draw attention. Respondents are instructed that "[f]or acts that [respondents] have not specifically done, [they need to] estimate [their] creative potential based on [their] performance on similar tasks" (Kaufman, 2012, p. 300). It is possible that respondents' estimations of their creative behaviors based on *similar tasks* may vary, and this situation may lower the stability of responses for the Self/Everyday subscale as well as the other subscales. That is because if an instrument's scoring is based on some type of rating and if raters (respondents in the case of the K-DOCS) are not well-informed on the use of the instrument, reliability of scores varies considerably from one respondent to another (American Educational Research Association et al., 2014).

Four categorical (publication type, culture, population, and focus of the study) and two continuous (publication year and female percentage) variables failed to reach statistical significance with respect to the reliability of K-DOCS total scores and subscale scores. These findings show the absence of systematic error and indicate that the scale does not favor any particular group in terms of item interpretation or response patterns with regard to those variables. The K-DOCS seems to produce equally reliable scores in different types of studies (e.g., published studies vs. unpublished studies), in different study settings (i.e., exploratory research vs. reliability or validity research), and in different cultures (i.e., Eastern vs. Western cultures).

The literature has evidence that shows similarities in creativity between Eastern and Western cultures. For instance, Chen et al. (2002) worked with European American ($n = 50$) and Chinese ($n = 48$) university students to compare creativity levels based on drawings of geometric shapes. The products were assessed

through the Consensual Assessment Technique. The drawings were found to be similar for both samples regarding creativity level. Guo et al. (2021) examined measurement invariance of divergent thinking tasks for culture. University students from China ($n = 316$) and the United States ($n = 302$) were the participants. Configural and metric invariance were supported for fluency and originality scores. Consistent with these results, our findings suggest that similarities between Eastern and Western cultures produce equally reliable K-DOCS scores. Therefore, the K-DOCS can be used for cross-cultural research.

It has been 10 years since the K-DOCS was developed. However, the scale still provides reliable estimates of individuals' creativity levels. Based on this finding, we can argue that the K-DOCS can be used for longitudinal studies. Finally, regardless of the female percentage in the study, the K-DOCS can be administered. This point is not surprising because measurement invariance was supported for gender for the K-DOCS subscale scores (Brauer et al., 2022; Kapoor et al., 2021). The items seem to work the same for female and male respondents with respect to giving stable responses on the perceptions of one's own creative behavior.

Variables with a significant impact

On the other hand, certain moderator variables significantly impacted the reliability of Scholarly, Performance, Scientific, and Artistic subscale scores. Before further discussing those variables, we want to emphasize that the statistically significant results obtained for those variables may not be due to systematic errors originating from certain variables; rather, the results may be due to unevenly distributed cells or merging certain subgroups (e.g., different countries) into a larger subgroup (e.g., Eastern countries). Thus, the differences should be interpreted with caution due to the low number of studies for certain moderator variables (e.g., test version and culture).

The North American, English-speaking, and U.S. samples always had the highest α coefficients compared to the coefficients estimated for their counterparts for the Scientific, Performance, and Artistic subscales. These results are not surprising considering that the K-DOCS was developed in the United States. However, we do not attribute the differences regarding these three variables to culture because culture did not have a statistically significant impact on the reliability of K-DOCS scores in our study. Rather, our findings indicate that the accuracy of K-DOCS scores is reduced during the adaptation process of the scale outside the United States. It is likely that some information is lost during the translation process and that this missing information lowers the reliability of K-DOCS scores outside the United States and in non-English-speaking countries.

Caucasian samples produced the lowest α coefficients for the Performance and Scientific subscales. This is an eye-catching finding because the K-DOCS was developed in the United States where the majority of the population is Caucasian. Nevertheless, the literature has evidence that is aligned with this result. For instance, in Paletz and Peng's (2009) study, the α coefficients estimated for the Caucasian participants (between .42 and .80) were lower than the α coefficients estimated for the Asian American participants (between .64 and .85).

There are some possible reasons for obtaining such a result for the Performance subscale. One reason may be related to the content of the items. Consider the following items: "Spontaneously creating lyrics to a rap song" and "Writing a poem" (Kaufman, 2012, p. 307). The literature indicates that rap music is more closely associated with African-American individuals than with Caucasian individuals (Charry, 2012; Elligan, 2000; Keyes, 1996). With respect to creative performance in literature, differences between Caucasian and other ethnicities were reported previously (Kaufman et al., 2004). Therefore, it is likely that some items in this subscale are more or less difficult for the Caucasian respondents and that the difficulty levels of the items lower the variability for samples with high Caucasian percentages (American Educational Research Association et al., 2014). It is possible that the samples with mixed ethnicities have higher variability due to varying levels of creative behavior and that the variation in mixed samples leads to more reliable scores.

With respect to the Scientific subscale, we cannot pinpoint any particular item content that leads to low reliability for Caucasian respondents. However, the literature shows some differences among ethnicities regarding engagement with as well as achievement in math and science (Lewis et al., 2009; Riegle-Crumb et al., 2011). These differences may be the reason for our findings. It is also possible that the items in these two subscales are more or less difficult for the Caucasian samples and that the difficulty levels of the items lower the variability (American Educational Research Association et al., 2014). Another reason may be that the Caucasian respondents are more familiar with the content of the items in these two subscales and that this familiarity lowers the variability (American Educational Research Association et al., 2014).

The educational levels of the samples impacted the α coefficients in a significant manner for the Scholarly and Performance subscales. There is no surprise in these results because creative performance is displayed at higher levels as years of education and age increase (Simonton, 1997). The variability among the respondents with higher levels of education seems to produce more consistent scores. Note that high school samples produced the lowest overall α values. The reason is possibly that the items in these two subscales are more difficult for high school students and that the mismatch between creative ability and item difficulty lowers the variability.

Furthermore, both the Scholarly and Performance subscales are not robust against the impact of training. Consider the following items: “Researching a topic using many different types of sources that may not be readily apparent,” “Gathering the best possible assortment of articles or papers to support a specific point of view,” and “Learning how to play a musical instrument” (Kaufman, 2012, p. 307). Although Kandemir and Kaufman (2020) found no statistically significant differences among majors for Scholarly subscale scores, our findings suggest that individuals exposed to education in a specific domain interpret the items more consistently and provide more stable responses for both the Scholarly and Performance subscales.

Finally, the test version has an impact on the α coefficients for the Scientific subscale. As expected, the original version provided the highest α coefficients (American Educational Research Association et al., 2014). Validated translations of the full version provided the second-highest coefficients. Interestingly, the 20-item version provided more reliable results than the 42-item version. However, the 20-item version was administered in English, while the 42-item version was administered in Turkish. It is likely that the accuracy of the items was reduced for the Scientific subscale during the Turkish adaptation, although the Turkish version included more items—note that the 20-item version has four and the 42-item version has 10 items. This finding makes more sense considering that English-speaking samples produced higher reliability coefficients compared to non-English-speaking samples. Nevertheless, it should be kept in mind that the numbers of studies that administered the 20-item and 42-item version were considerably low. If there were more studies on the 20-item and 42-item versions, the results might be different.

In terms of the continuous variables, the percentage of Caucasians had a negative impact on the α coefficients for the Performance subscale. This finding is aligned with the finding obtained for ethnicity. Note that the Caucasian samples produced the least reliable scores for the Performance subscale.

As the mean and *SD* of the pooled sample age increased, so did the α coefficients for the Scholarly, Performance, Scientific and Artistic subscales. These results are not surprising considering that cognitive development and creative performance is expected to increase with age (Gros Lambert & Mahon, 2006; Piaget & Inhelder, 1966; Simonton, 1997). Note that this finding is aligned with the findings obtained for the variable educational level.

RELIABILITY INDUCTION

In order to answer the third research question, which addresses reliability induction, we estimated the reliability induction rates and compared the inducing studies with the reporting studies for certain variables. The induction rate estimated in our study (39.62%) is lower than the induction rate reported by Sánchez-Meca et al. (78.6%; 2021). Nevertheless, our findings show that considerable data are still not available for estimating the overall α coefficients of the K-DOCS and examining the effects of the moderator variables. This situation poses substantial risk because when several studies commit reliability induction, the pooled reliability coefficient is estimated with a large amount of error (Sánchez-Meca et al., 2021). This situation should be kept in mind when interpreting our findings.

Furthermore, reliability induction may impact the refinement of the K-DOCS. The reason is that reliability induction may have led to an overly optimistic view of the reliability of K-DOCS scores. When there is missing information on reliability and when reliability is considered a property that remains constant throughout all administrations of an instrument, researchers miss the opportunity to identify and address potential sources of error in scores on the instrument (Sánchez-Meca et al., 2021).

It is noteworthy that our study found no statistical significance between some characteristics of inducing and reporting studies, indicating that reliability induction does not threaten the validity of our findings (Sánchez-Meca et al., 2021). However, this should not diminish the importance of reporting reliability, as the reliability of scores on an instrument varies across samples (American Educational Research Association et al., 2014). Therefore, it is imperative that researchers using the K-DOCS should estimate and report the reliability coefficients for each application. This practice is not only vital for the K-DOCS’s credibility but also for enhancing the reliability and validity of scale scores.

CONCLUSIONS, LIMITATIONS, AND FUTURE DIRECTIONS

The present study is the first one that examines the overall reliability coefficients, the variability of the reliability of K-DOCS scores, and reliability induction for the K-DOCS. Our findings show that the K-DOCS produces highly reliable total scores and subscale scores. The reliability of scores on the K-DOCS seems to be resistant to different variables, including publication type, culture, population, focus of the study, publication year, and female percentage. However, a few other variables (e.g., ethnicity, test version, and mean age) may affect the reliability of certain subscale scores. Researchers and practitioners should be cautious about the effects of those variables when using the K-DOCS and focusing on the Scholarly, Performance, Scientific, and Artistic subscales.

This study has some limitations. The first is the type of reliability coefficient considered for the analyses. We conducted the RG analyses with the α coefficients because we could not conduct RG analyses for test-retest reliability and omega coefficients due to the number of coefficients. The second limitation is the number of included studies. This limitation is a result of both reliability induction and the criterion for language—note that only the studies reported in English were kept. Another limitation is the missing data for certain moderator variables—note that some studies failed to report the values of certain variables. Finally, creating subgroups for each categorical variable is also a limitation. The reason is that certain information may have been lost when the subgroups were generated.

Future RG meta-analysis studies are needed to examine the K-DOCS. Future studies should be conducted with other transformation methods and multivariate models. Additionally, after new studies are published, RG meta-analyses should be conducted to investigate the impact of other moderator variables. Moreover, when the number of studies on each short version of the scale is sufficient, the 20-item and 42-item versions should be examined separately. Finally, when there are a sufficient number of studies that report a reliability coefficient other than α , that type of reliability coefficient should be examined through the RG approach.

At this point, we want to emphasize that reporting omega coefficients will be a better practice for the K-DOCS in future studies. This is because the K-DOCS items are possibly not tau-equivalent, the response options are on an ordinal scale, and the scale is multidimensional. Note that the omega coefficient provides better estimates when tau-equivalency is violated (Deng & Chan, 2017). Furthermore, there is a version of omega to estimate a reliability coefficient for multidimensional instruments (Şimşek & Noyan, 2013), and the omega coefficient yields more precise reliability estimates when the data are in an ordinal scale (McNeish, 2018).

CONFLICT OF INTEREST STATEMENT

We have no known conflict of interest to disclose.

DATA AVAILABILITY STATEMENT

Materials and analysis code for this study are available by emailing the corresponding author.

REFERENCES

- Note: References with an asterisk (*) denote the studies included in the RG meta-analyses (Table S1).
- *Akar, I., & Coskun, B.K. (2020). Exploring the relationship between creativity and cyberloafing of prospective teachers. *Thinking Skills and Creativity*, 38, 100724; doi: [10.1016/j.tsc.2020.100724](https://doi.org/10.1016/j.tsc.2020.100724).
 - *Altınay, L., Kinali-Madanoglu, G., Kromidha, E., Nurmagambetova, A., & Madanoglu, M. (2021). Mental aspects of cultural intelligence and self-creativity of nascent entrepreneurs: The mediating role of emotionality. *Journal of Business Research*, 131, 793–802.
 - *Altınay, L., Kromidha, E., Nurmagambetova, A., Alrawdiah, Z., & Madanoglu, G.K. (2022). A social cognition perspective on entrepreneurial personality traits and intentions to start a business: Does creativity matter? *Management Decision*, 60, 1606–1625; doi: [10.1108/MD-12-2020-1592](https://doi.org/10.1108/MD-12-2020-1592).
 - Amabile, T.M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology*, 43, 997–1013; doi: [10.1037/0022-3514.43.5.997](https://doi.org/10.1037/0022-3514.43.5.997).
 - American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
 - Ashraf, F., Nusrat, A., & Mehboob, D. (2019). Predictive role of personality characteristics in positive emotions and creativity: A study of Pakistani nuns. *Pakistan Journal of Psychological Research*, 34, 383–400; doi: [10.33824/pjpr.2019.34.2.21](https://doi.org/10.33824/pjpr.2019.34.2.21).
 - *Awawdeh, M.N.A., & Lian, L.H. (2020). The relationship between creativity domains and academic environment. *International Journal of Innovation, Creativity and Change*, 13, 928–940.

- *Awofala, A.O.A., & Fatade, A.O. (2015). Validation of the domains of creativity scale for Nigerian preservice science, technology, and mathematics teachers. *Electronic Journal of Research in Educational Psychology*, 13(1), 131–150; doi: [10.14204/ejrep.35.14057](https://doi.org/10.14204/ejrep.35.14057).
- Aznar, A., Sowden, P., Bayless, S., Ross, K., Warhurst, A., & Pachi, D. (2021). Home-schooling during COVID-19 lockdown: Effects of coping style, home space, and everyday creativity on stress and home-schooling outcomes. *Couple and Family Psychology: Research and Practice*, 10, 294–312; doi: [10.1037/cfp0000182](https://doi.org/10.1037/cfp0000182).
- Baer, J. (1998). The case for domain specificity of creativity. *Creativity Research Journal*, 11, 173–177; doi: [10.1207/s15326934crj1102_7](https://doi.org/10.1207/s15326934crj1102_7).
- Baer, J. (2012). Domain specificity and the limits of creativity theory. *The Journal of Creative Behavior*, 46(1), 16–29; doi: [10.1002/jocb.002](https://doi.org/10.1002/jocb.002).
- Baer, J. (2015). The importance of domain-specific expertise in creativity. *Roeper Review*, 37, 165–178; doi: [10.1080/02783193.2015.1047480](https://doi.org/10.1080/02783193.2015.1047480).
- Baer, J., & Kaufman, J.C. (2005). Bridging generality and specificity: The amusement park theoretical (APT) model of creativity. *Roeper Review*, 27(3), 158–163; doi: [10.1080/02783190509554310](https://doi.org/10.1080/02783190509554310).
- Begg, C.B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 50, 1088–1101; doi: [10.2307/2533446](https://doi.org/10.2307/2533446).
- Beghetto, R.A. (2006). Creative self-efficacy: Correlates in middle and secondary students. *Creativity Research Journal*, 18, 447–457; doi: [10.1207/s15326934crj1804_4](https://doi.org/10.1207/s15326934crj1804_4).
- *Benić, M.Ž. (2021). Translation and validation of the Kaufman domains of creativity scale on a Croatian sample of early childhood and preschool education students. *Center for Educational Policy Studies Journal*, 11, 163–179; doi: [10.26529/cepsj.708](https://doi.org/10.26529/cepsj.708).
- *Bertašienė, I. (2020). Understanding future teachers' perceptions of their creativity. *Contemporary Research on Organization Management and Administration*, 8(2), 107–119; doi: [10.33605/croma-022020-007](https://doi.org/10.33605/croma-022020-007).
- Boccia, M., Piccardi, L., Palermo, L., Nori, R., & Palmiero, M. (2015). Where do bright ideas occur in our brain? Meta-analytic evidence from neuroimaging studies of domain-specific creativity. *Frontiers in Psychology*, 6, 1195; doi: [10.3389/fpsyg.2015.01195](https://doi.org/10.3389/fpsyg.2015.01195).
- Bonett, D.G. (2002a). Sample size requirements for testing and estimating coefficient alpha. *Journal of Educational and Behavioral Statistics*, 27, 335–340.
- Bonett, D. G. (2002b). Sample size requirements for estimating intraclass correlations with desired precision. *Statistics in Medicine*, 21(9), 1331–1335; doi: [10.1002/sim.1108](https://doi.org/10.1002/sim.1108)
- *Boot, N., Neuvicka, B., & Baas, M. (2020). Creativity in ADHD: Goal-directed motivation and domain specificity. *Journal of Attention Disorders*, 24(13), 1857–1866; doi: [10.1177/1087054717727352](https://doi.org/10.1177/1087054717727352).
- Botella, J., Suero, M., & Gamba, H. (2010). Psychometric inferences from a meta-analysis of reliability and internal consistency coefficients. *Psychological Methods*, 4, 386–397.
- Brauer, K., Sendatzki, R., Kaufman, J.C., & Proyer, R.T. (2022). Counting the muses in German speakers—Evaluation of the German-language translation of the Kaufman domains of creativity scales (K-DOCS). *Psychological Test Adaptation and Development*, 3(1), 70–84; doi: [10.1027/2698-1866/a000024](https://doi.org/10.1027/2698-1866/a000024).
- *Cajilog, C.F.S. (2018). *The enculturation scale for Filipino-Americans: A psychometric investigation and its relationship to attitudes toward help seeking* [Unpublished doctoral dissertation]. Berkeley, CA: Wright Institute Graduate School.
- Carson, S.H., Peterson, J.B., & Higgins, D.M. (2005). Reliability, validity, and factor structure of the creative achievement questionnaire. *Creativity Research Journal*, 17, 37–50; doi: [10.1207/s15326934crj1701_4](https://doi.org/10.1207/s15326934crj1701_4).
- Charry, E. (Ed.). (2012). *Hip hop Africa: New African music in a globalizing world*. Bloomington, IN: Indiana University Press. <http://www.jstor.org/stable/j.ctt16gz8kj>.
- Chen, C., Himsel, A., Kasof, J., Greenberger, E., & Dmitrieva, J. (2006). Boundless creativity: Evidence for the domain generality of individual differences in creativity. *The Journal of Creative Behavior*, 40, 179–199; doi: [10.1002/j.2162-6057.2006.tb01272.x](https://doi.org/10.1002/j.2162-6057.2006.tb01272.x).
- 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, 171–187.
- Cochran, W.G. (1954). The combination of estimates from different experiments. *Biometrics*, 10, 101–129.
- *Collier, W.G. (2022). Future time perspective and self-ratings of creativity. *ATSK Journal of Psychology*, 3(1), 17–24.
- *Cotas-Girard, A., Cajillog, C.F., Del Prado, A., & McGovern, K. (2022). A psychometric investigation of the enculturation scale for Filipino Americans-short form: Further reliability, validation, and confirmation of factor structure. *Asian American Journal of Psychology*, 13(1), 30–40; doi: [10.1037/aap0000245](https://doi.org/10.1037/aap0000245).
- Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334; doi: [10.1007/bf02310555](https://doi.org/10.1007/bf02310555).
- *De Fazio, C., Dennis, T., English, R., & Brent Sattelmeyer, B. (2019). *Creative coping strategies: Reducing academic stress in college students* [Unpublished master's thesis]. Gainesville, GA: Brenau University.
- Deng, J., Wang, M.-C., Zhang, X., Shou, Y., Gao, Y., & Luo, J. (2019). The inventory of callous unemotional traits: A reliability generalization meta-analysis. *Psychological Assessment*, 31(6), 765–780; doi: [10.1037/pas0000698](https://doi.org/10.1037/pas0000698).
- Deng, L., & Chan, W. (2017). Testing the difference between reliability coefficients alpha and omega. *Educational and Psychological Measurement*, 77, 185–203; doi: [10.1177/0013164416658325](https://doi.org/10.1177/0013164416658325).
- Dollinger, S.J. (2003). Need for uniqueness, need for cognition, and creativity. *The Journal of Creative Behavior*, 37(2), 99–116; doi: [10.1002/j.2162-6057.2003.tb00828.x](https://doi.org/10.1002/j.2162-6057.2003.tb00828.x).
- Dong, Y., Lin, J., Li, H., Cheng, L., Niu, W., & Tong, Z. (2022). How parenting styles affect children's creativity: Through the lens of self. *Thinking Skills and Creativity*, 45, 101045.

AN RG META-ANALYSIS OF THE K-DOCS

- *Dostál, D., Plháková, A., & Zášková, T. (2017). Domain-specific creativity in relation to the level of empathy and systemizing. *The Journal of Creative Behavior*, 51, 225–239.
- Dousay, T., & Weible, A. (2019). Build-a-bug workshop: Designing a learning experience with emerging technology to foster creativity. *TechTrends*, 63(1), 41–52.
- Dow, G.T., & Mayer, R.E. (2004). Teaching students to solve insight problems: Evidence for domain specificity in creativity training. *Creativity Research Journal*, 16, 389–398; doi: [10.1080/10400410409534550](https://doi.org/10.1080/10400410409534550).
- *Dudău, D.P. (2013). *The impostor phenomenon and its relation to creative potential among high school students: Could fraudulence feelings regarding creativity be justified* [Conference session]? Bucharest, Romania: The International Conference ‘Education and Creativity for a Knowledge-Based Society’.
- Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56, 455–463; doi: [10.1111/j.0006-341X.2000.00455.x](https://doi.org/10.1111/j.0006-341X.2000.00455.x).
- Egger, M., Smith, G.D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315, 629–634.
- Elligan, D. (2000). Rap therapy: A culturally sensitive approach to psychotherapy with young African American men. *Journal of African American Men*, 5(3), 27–36. <https://www.jstor.org/stable/41819404>.
- Faggion, C.M., Jr., Menne, M.C., & Pandis, N. (2021). Prediction intervals should be included in meta-analyses published in dentistry. *European Journal of Oral Sciences*, 129(6), e12827; doi: [10.1111/eos.12827](https://doi.org/10.1111/eos.12827).
- Feist, G.J. (2019a). The function of personality in creativity: Updates on the creative personality. In J.C. Kaufman & R.J. Sternberg (Eds.), *The Cambridge handbook of creativity* (pp. 353–373). Cambridge: Cambridge University Press. doi: [10.1017/9781316979839.019](https://doi.org/10.1017/9781316979839.019).
- Feist, G.J. (2019b). Creativity and the big two model of personality: Plasticity and stability. *Current Opinion in Behavioral Sciences*, 27, 31–35.
- *Fisher, J., & Davis, L. (2018). *Stress and self-efficacy: The relationship with everyday creativity* [Unpublished master’s thesis]. Gainesville, GA: Brenau University.
- Funke, J. (2009). On the psychology of creativity. In P. Meusburger, J. Funke, & E. Wunder (Eds.), *Milieus of creativity: An interdisciplinary approach to spatiality of creativity* (pp. 11–23). Heidelberg: Springer; doi: [10.1007/978-1-4020-9877-2_2](https://doi.org/10.1007/978-1-4020-9877-2_2).
- Graham, J.M., Diebels, K.J., & Barnow, Z.B. (2011). The reliability of relationship satisfaction: A reliability generalization meta-analysis. *Journal of Family Psychology*, 25(1), 39–48; doi: [10.1037/a0022441](https://doi.org/10.1037/a0022441).
- Graham, J.M., Liu, Y.J., & Jeziorski, J.L. (2006). The dyadic adjustment scale: A reliability generalization meta-analysis. *Journal of Marriage and Family*, 68, 701–717; doi: [10.1111/j.1741-3737.2006.00284.x](https://doi.org/10.1111/j.1741-3737.2006.00284.x).
- Grajzel, K., Acar, S., & Singer, G. (2023). The Big Five and divergent thinking: A meta-analysis. *Personality and Individual Differences*, 214, 112338; doi: [10.1016/j.paid.2023.112338](https://doi.org/10.1016/j.paid.2023.112338).
- Gros Lambert, A., & Mahon, A.D. (2006). Perceived exertion: influence of age and cognitive development. *Sports Medicine*, 36, 911–928.
- Guilford, J.P. (1966). Measurement and creativity. *Theory Into Practice*, 5(4), 185–189.
- Guo, Y., Lin, S., Guo, J., Lu, Z.L., & Shangguan, C. (2021). Cross-cultural measurement invariance of divergent thinking measures. *Thinking Skills and Creativity*, 41, 100852; doi: [10.1016/j.tsc.2021.100852](https://doi.org/10.1016/j.tsc.2021.100852).
- *Hanakis, M.C. (2017). *Problem solving style and creative achievement* [Unpublished doctoral dissertation]. New York, NY: Fordham University.
- Henson, R.K., & Thompson, B. (2002). Characterizing measurement error in scores across studies: Some recommendations for conducting “reliability generalization” studies. *Measurement and Evaluation in Counseling and Development*, 35, 113–126.
- Higgins, J.P., Thompson, S.G., Deeks, J.J., & Altman, D.G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, 327, 557–560.
- Hinton, B. L. (1968). A model for the study of creative problem solving. *The Journal of Creative Behavior*, 2, 133–142. [10.1002/j.2162-6057.1968.tb00095.x](https://doi.org/10.1002/j.2162-6057.1968.tb00095.x)
- Hocevar, D. (1979). *The development of the Creative Behavior Inventory (CBI)*. Denver, CO: Paper presented at the annual meeting of the Rocky Mountain Psychological Association (ERIC Document Reproduction Service No.Ed. 170 350).
- *Hong, O., Park, M.H., & Song, J. (2022). The assessment of science classroom creativity: Scale development. *International Journal of Science Education*, 44, 1356–1377; doi: [10.1080/09500693.2022.2077466](https://doi.org/10.1080/09500693.2022.2077466).
- *Jonason, P.K., Richardson, E.N., & Potter, L. (2015). Self-reported creative ability and the dark triad traits: An exploratory study. *Psychology of Aesthetics, Creativity, and the Arts*, 9, 488–494; doi: [10.1037/aca0000037](https://doi.org/10.1037/aca0000037).
- Jung, M., Kim, H.S., Loprinzi, P.D., & Kang, M. (2021). Serial-multiple mediation of enjoyment and intention on the relationship between creativity and physical activity. *AIMS Neuroscience*, 8(1), 161–180; doi: [10.3934/Neuroscience.2021008](https://doi.org/10.3934/Neuroscience.2021008).
- Kandemir, M.A., & Kaufman, J.C. (2020). The kaufman domains of creativity scale (K-DOCS): Turkish validation and relationship to academic major. *The Journal of Creative Behavior*, 54, 1002–1012; doi: [10.1002/jocb.428](https://doi.org/10.1002/jocb.428).
- *Kapoor, H., & Kaufman, J.C. (2021). Are cheaters common or creative? Person-situation interactions of resistance in learning contexts. *Journal of Academic Ethics*, 19, 157–174; doi: [10.1007/s10805-020-09379-w](https://doi.org/10.1007/s10805-020-09379-w).
- Kapoor, H., Reiter-Palmon, R., & Kaufman, J.C. (2021). Norming the muses: Establishing the psychometric properties of the Kaufman domains of creativity scale. *Journal of Psychoeducational Assessment*, 39, 680–693; doi: [10.1177/07342829211008334](https://doi.org/10.1177/07342829211008334).

- Kapoor, H., Zheng, M., Reiter-Palmon, R., & Kaufman, J.C. (2023). Toward equitable creativity self-assessment: Measurement invariance and mean differences across ethnicity in the Kaufman Domains of Creativity Scale. *Psychology of Aesthetics, Creativity, and the Arts*, 17, 487–494; doi: [10.1037/aca0000594](https://doi.org/10.1037/aca0000594).
- *Kaufman, J.C. (2012). Counting the muses: Development of the kaufman domains of creativity scale (K-DOCS). *Psychology of Aesthetics, Creativity, and the Arts*, 6, 298–308; doi: [10.1037/a0029751](https://doi.org/10.1037/a0029751).
- Kaufman, J.C., Baer, J., & Gentile, C.A. (2004). Differences in gender and ethnicity as measured by ratings of three writing tasks. *The Journal of Creative Behavior*, 39, 56–69; doi: [10.1002/j.2162-6057.2004.tb01231.x](https://doi.org/10.1002/j.2162-6057.2004.tb01231.x).
- Kaufman, J.C., Cole, J.C., & Baer, J. (2009). The construct of creativity: A structural model for self-reported creativity ratings. *The Journal of Creative Behavior*, 43, 119–134; doi: [10.1002/j.2162-6057.2009.tb01310.x](https://doi.org/10.1002/j.2162-6057.2009.tb01310.x).
- *Kaufman, J.C., Kapoor, H., Patston, T., & Croypley, D.H. (2021). Explaining standardized educational test scores: The role of creativity above and beyond GPA and personality. *Psychology of Aesthetics, Creativity, and the Arts*. Advance online publication; doi: [10.1037/aca0000433](https://doi.org/10.1037/aca0000433).
- Keyes, C.L. (1996). At the crossroads: Rap music and its African nexus. *Ethnomusicology*, 40, 223–248. <https://www.jstor.org/stable/852060>.
- *Kibici, V.B. (2022). An analysis of the relationships between secondary school students' creativity, music achievement and attitudes. *International Journal on Social and Education Sciences*, 4(1), 87–100; doi: [10.46328/ijones.304](https://doi.org/10.46328/ijones.304).
- King, L.A., Walker, L.M., & Broyles, S.J. (1996). Creativity and the five-factor model. *Journal of Research in Personality*, 30, 189–203; doi: [10.1006/jrpe.1996.0013](https://doi.org/10.1006/jrpe.1996.0013).
- *Kircaburun, K., Griffiths, M.D., Şahin, F., Bahtiyar, M., Atmaca, T., & Tosuntaş, Ş.B. (2020). The mediating role of self/everyday creativity and depression on the relationship between creative personality traits and problematic social media use among emerging adults. *International Journal of Mental Health and Addiction*, 18, 77–88; doi: [10.1007/s11469-018-9938-0](https://doi.org/10.1007/s11469-018-9938-0).
- *Komsu, U.C. (2021). Postgraduate students' perceptions of research self-efficacy and critical thinking disposition and their impact on academic creativity: Case of Mersin University. *Turkish Online Journal of Educational Technology*, 20, 53–66.
- *Lee, A.W. (2018). *Self-perceptions of creativity and creative performance in adolescents* [Unpublished master's thesis]. Cleveland, OH: Case Western Reserve University.
- *Lee, J.H. (2017). Building creative confidence through an interdisciplinary creativity course: Changes in creative challenges and creative personal identity. *Innovations in Education and Teaching International*, 59, 316–325; doi: [10.1080/14703297.2020.1835689](https://doi.org/10.1080/14703297.2020.1835689).
- *Lee, J.H., & Portillo, M. (2022). Transferability of creative self-belief across domains: The differential effects of a creativity course for university students. *Thinking Skills and Creativity*, 43, 100996; doi: [10.1016/j.tsc.2021.100996](https://doi.org/10.1016/j.tsc.2021.100996).
- Lewis, J., Menzies, H., Najera, E., & Page, R. (2009). Rethinking trends in minority participation in the sciences. *Science Education*, 93, 961–977; doi: [10.1002/sce.20338](https://doi.org/10.1002/sce.20338).
- Magnusson, M. (2018). *The creative networker predicted relations between network behavior and creativity* [Unpublished doctoral dissertation]. Uppsala: Uppsala University.
- *McKay, A.S., & Gutworth, M.B. (2021). Temporal individual differences and creativity: An exploratory investigation. *Psychology of Aesthetics, Creativity, and the Arts*, 15, 295–312; doi: [10.1037/aca0000292](https://doi.org/10.1037/aca0000292).
- *McKay, A.S., Karwowski, M., & Kaufman, J.C. (2017). Measuring the muses: Validating the Kaufman domains of creativity scale (K-DOCS). *Psychology of Aesthetics, Creativity, and the Arts*, 11, 216–230; doi: [10.1037/aca0000074](https://doi.org/10.1037/aca0000074).
- McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, 23, 412–433; doi: [10.1037/met0000144](https://doi.org/10.1037/met0000144).
- *Minney, J.A. (2016). *Isolation: The optimum environment for creativity? The relationship between the experience of ostracism and creativity* [Unpublished doctoral dissertation]. Tuscaloosa, AL: The University of Alabama.
- Miroshnik, K.G., Shcherbakova, O.V., & Kaufman, J.C. (2022). Kaufman domains of creativity scale: Relationship to occupation and measurement invariance across gender. *Creativity Research Journal*, 34, 159–177; doi: [10.1080/10400419.2021.1953823](https://doi.org/10.1080/10400419.2021.1953823).
- *Muhammad, K., Nadeem, D.M., & Nadeem, A. (2021). Relationship between psychoticism and creativity. *Journal of Professional & Applied Psychology*, 2, 199–205; doi: [10.52053/jpap.v2i2.37](https://doi.org/10.52053/jpap.v2i2.37).
- Mumford, M.D. (2003). Taking stock in taking stock. *Creativity Research Journal*, 15, 147–151; doi: [10.1080/10400419.2003.9651408](https://doi.org/10.1080/10400419.2003.9651408).
- *Mussel, P., McKay, A.S., Ziegler, M., Hewig, J., & Kaufman, J.C. (2015). Predicting creativity based on the facets of the theoretical intellect framework. *European Journal of Personality*, 29, 459–467; doi: [10.1002/per.2000](https://doi.org/10.1002/per.2000).
- *Mvududu, M.J. (2020). *Can trait emotional intelligence variables of well-being, self-control, emotionality, and sociability individually or collectively predict a software development engineer's creativity* [Unpublished doctoral dissertation]? Newberg, OR: George Fox University.
- Nagashima, K., Noma, H., & Furukawa, T.A. (2019). Prediction intervals for random-effects meta-analysis: a confidence distribution approach. *Statistical Methods in Medical Research*, 28, 1689–1702; doi: [10.1177/0962280218773520](https://doi.org/10.1177/0962280218773520).
- Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric Theory* (3rd edn). New York, NY: McGraw-Hill.
- O'Rourke, N., Hatcher, L., & Stepanski, E.J. (2005). *A step-by-step approach to using SAS for univariate & multivariate statistics*. Cary, NC: SAS Institute.
- *Osamika, B.E., & Usheninte, D. (2021). Influence of psychological distress and gender on creative thinking ability of students in tertiary institutions in Ibadan, Oyo State Nigeria. *Journal Of Capital Development In Behavioural Sciences*, 9, 64–87.
- Paletz, S.B.F., & Peng, K. (2009). Problem finding and contradiction: Examining the relationship between naive dialectical thinking, ethnicity, and creativity. *Creativity Research Journal*, 21(2–3), 139–151; doi: [10.1080/10400410902858683](https://doi.org/10.1080/10400410902858683).

AN RG META-ANALYSIS OF THE K-DOCS

- *Penagos-Corzo, J.C., & Saucedo, A. (2021). Development and validation of a questionnaire to assess creative potential traits. *Frontiers in Psychology*, 12, 756079; doi: [10.3389/fpsyg.2021.756079](https://doi.org/10.3389/fpsyg.2021.756079).
- Piaget, J., & Inhelder, B. (1966). *The psychology of the child*. New York, NY: Harper and Row.
- Plucker, J.A. (1998). Beware of simple conclusions: The case for content generality of creativity. *Creativity Research Journal*, 11, 179–182; doi: [10.1207/s15326934crj1102_8](https://doi.org/10.1207/s15326934crj1102_8).
- Plucker, J.A. (1999). Reanalyses of student responses to creativity checklists: Evidence of content generality. *The Journal of Creative Behavior*, 33, 126–137; doi: [10.1002/j.2162-6057.1999.tb01042.x](https://doi.org/10.1002/j.2162-6057.1999.tb01042.x).
- Plucker, J.A. (2004). Generalization of creativity across domains: Examination of the method effect hypothesis. *The Journal of Creative Behavior*, 38, 1–12; doi: [10.1002/j.2162-6057.2004.tb01228.x](https://doi.org/10.1002/j.2162-6057.2004.tb01228.x).
- Plucker, J.A., & Beghetto, R.A. (2004). Why creativity is domain general, why it looks domain specific, and why the distinction does not matter. In R.J. Sternberg, E.L. Grigorenko, & J.L. Singer (Eds.), *Creativity: From potential to realization* (pp. 153–167). Washington, DC: American Psychological Association; doi: [10.1037/10692-009](https://doi.org/10.1037/10692-009).
- *Pretz, J.E., & Kaufman, J.C. (2017). Do traditional admissions criteria reflect applicant creativity? *The Journal of Creative Behavior*, 51, 240–251; doi: [10.1002/jocb.120](https://doi.org/10.1002/jocb.120).
- *Raymond, E., Uduafemhe, E.M., Alome, S.A., & Ogumah, B.A.O. (2018). Correlation between the five personality dimensions and creative behaviours of electrical and electronics teachers in Nigerian technical colleges. *Journal of Technical Education and Training*, 10(2), 28–37.
- Riegler-Crumb, C., Moore, C., & Ramos-Wada, A. (2011). Who wants to have a career in science or math? Exploring adolescents' future aspirations by gender and race/ethnicity. *Science Education*, 95, 458–476; doi: [10.1002/sce.20431](https://doi.org/10.1002/sce.20431).
- *Royston, R.P. (2016). *The relationship between Big-C, little-c, and Pro-c creativity and fixed and malleable creative mindsets* [Unpublished master's thesis]. Omaha, NE: University of Nebraska at Omaha.
- Runco, M.A. (2007). *Creativity: Theories and themes: Research, development, and practice*. Boston, MA: Elsevier Academic Press.
- Runco, M.A., & Jaeger, G.J. (2012). The standard definition of creativity. *Creativity Research Journal*, 24, 92–96; doi: [10.1080/10400419.2012.650092](https://doi.org/10.1080/10400419.2012.650092).
- Runco, M.A., Plucker, J.A., & Lim, W. (2001). Development and psychometric integrity of a measure of ideational behavior. *Creativity Research Journal*, 13, 393–400; doi: [10.1207/S15326934CRJ1334_16](https://doi.org/10.1207/S15326934CRJ1334_16).
- *Şahin, F. (2016a). Kaufman alanları yaratıcılık ölçeği'nin Türkçe'ye uyarlanması ve psikometrik özelliklerinin incelenmesi [Adaptation of the Kaufman domains of creativity scale into Turkish and examination of its psychometric properties]. *Elementary Education Online*, 15, 855–867; doi: [10.17051/ee.2016.70479](https://doi.org/10.17051/ee.2016.70479).
- *Şahin, F. (2016b). General intelligence, emotional intelligence and academic knowledge as predictors of creativity domains: A study of gifted students. *Cogent Education*, 3, 1218315; doi: [10.1080/2331186X.2016.1218315](https://doi.org/10.1080/2331186X.2016.1218315).
- *Şahin, F. (2017). Chess: A game of kings or the king of games? A Study of creativity in gifted and non-gifted students. *Karaelmas Eğitim Bilimleri Dergisi*, 5, 263–272.
- *Şahin, F., Özer, E., & Deniz, M.E. (2016). The predictive level of emotional intelligence for the domain-specific creativity: A study on gifted students. *Eğitim ve Bilim*, 41, 181–197.
- 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, 516–536; doi: [10.1002/jrsm.1487](https://doi.org/10.1002/jrsm.1487).
- *Saw, K.N.N., & Han, B. (2022). Prospective teachers' academic success in Myanmar: Successful intelligence and personality effects. *Journal of Education for Teaching*, 48, 72–88; doi: [10.1080/02607476.2021.1959266](https://doi.org/10.1080/02607476.2021.1959266).
- 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, 139–148; doi: [10.1037/a0014940](https://doi.org/10.1037/a0014940).
- Simonton, D.K. (1997). Creative productivity: A predictive and explanatory model of career trajectories and landmarks. *Psychological Review*, 104, 66–89; doi: [10.1037/0033-295X.104.1.66](https://doi.org/10.1037/0033-295X.104.1.66).
- Simonton, D.K. (2017). Domain-general creativity: On generating original, useful, and surprising combinations. In J.C. Kaufman, V.P. Glăveanu, & J. Baer (Eds.), *The Cambridge handbook of creativity across domains* (pp. 41–60). Cambridge: Cambridge University Press; doi: [10.1017/9781316274385.004](https://doi.org/10.1017/9781316274385.004).
- Şimşek, G., & Noyan, F. (2013). McDonald's ω , Cronbach's α , and generalized θ for composite reliability of common factors structures. *Communications in Statistics - Simulation and Computation*, 42, 2008–2025; doi: [10.1080/03610918.2012.689062](https://doi.org/10.1080/03610918.2012.689062).
- *Soles, K.L. (2020). *Relationship between the big five and creative self-beliefs in undergraduates in terms of college enrollment and major* [Unpublished doctoral dissertation]. Texas, TX: University of North Texas.
- Starko, A.J. (2014). *Creativity in the classroom: Schools of curious delight*. New York: Routledge.
- Stein, M.I. (1953). Creativity and culture. *Journal of Psychology*, 36, 311–322.
- Sternberg, R.J. (2009). Domain-generality versus domain-specificity of creativity. In P. Meusburger, J. Funke, & E. Wunder (Eds.), *Milieux of creativity: An interdisciplinary approach to spatiality of creativity* (pp. 25–38). New York, NY: Springer; doi: [10.1007/978-1-4020-9877-2_3](https://doi.org/10.1007/978-1-4020-9877-2_3).
- Susanto, E., Novitasari, Y., Ahman, A., Nurhuda, N., Rakhmat, C., Hidayat, A., & Wibowo, S.B. (2018). Rasch model analysis of Kaufman domains of creativity scale (K-DOCS) to Indonesian students. *Journal of Physics: Conference Series*, 1114(1), 1–7; doi: [10.1088/1742-6596/1114/1/012027](https://doi.org/10.1088/1742-6596/1114/1/012027).
- *Tan, C.S., Tan, S.A., Cheng, S.M., Hashim, I.H.M., & Ong, A.W.H. (2021). Development and preliminary validation of the 20-item Kaufman domains of creativity scale for use with Malaysian populations. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues*, 40, 1946–1957; doi: [10.1007/s12144-019-0124-8](https://doi.org/10.1007/s12144-019-0124-8).

- Tan, C.S., Tan, S.A., May, C.S., & Hashim, I.H.M. (2016). Factor structure and psychometric qualities of the Kaufman domains of creativity scale [Conference session]. In *International Conference on Education and Psychology*. Sabah, Malaysia.
- *Tan, M., Mourguet, C., Hein, S., MacCormick, J., Barbot, B., & Grigorenko, E. (2015). Differences in judgments of creativity: How do academic domain, personality, and self-reported creativity influence novice judges' evaluations of creative productions? *Journal of Intelligence*, 3(3), 73–90; doi: [10.3390/jintelligence303007](https://doi.org/10.3390/jintelligence303007).
- *Taylor, C.L., & Kaufman, J.C. (2021). The creative trait motivation scales. *Thinking Skills and Creativity*, 39, 100763; doi: [10.1016/j.tsc.2020.100763](https://doi.org/10.1016/j.tsc.2020.100763).
- Thompson, B., & Vacha-Haase, T. (2000). Psychometrics is datametrics: The test is not reliable. *Educational and Psychological Measurement*, 60, 174–195; doi: [10.1177/00131640021970448](https://doi.org/10.1177/00131640021970448).
- *Tidikis, V., & Dunbar, N.D. (2019). Openness to experience and creativity: When does global citizenship matter? *International Journal of Psychology*, 54, 264–268; doi: [10.1002/ijop.12463](https://doi.org/10.1002/ijop.12463).
- *Top, E., & Akil, M. (2018). Effects of the sports on the personality traits and the domains of creativity. *World Journal of Education*, 8, 56–64; doi: [10.5430/WJE.V8N3P56](https://doi.org/10.5430/WJE.V8N3P56).
- Torrance, E.P. (1974). *Torrance tests of creative thinking*. Princeton, NJ: Personnel Press.
- *Tu, C., Dille, A.E., & Kaufman, J.C. (2015). Do we create what we watch? Creativity and entertainment preferences. *Psychology of Aesthetics, Creativity, and the Arts*, 9, 394–404.
- *Tu, C., Guo, J., Hatcher, R.C., & Kaufman, J.C. (2018). The relationship between emotional intelligence and domain-specific and domain-general creativity. *The Journal of Creative Behavior*, 54, 337–349; doi: [10.1002/jocb.369](https://doi.org/10.1002/jocb.369).
- 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](https://doi.org/10.1177/0013164498058001002).
- Van den Noortgate, W., López-López, J.A., Marín-Martínez, F., & Sánchez-Meca, J. (2015). Meta-analysis of multiple outcomes: A multilevel approach. *Behavior Research Methods*, 47, 1274–1294; doi: [10.3758/s13428-014-0527-2](https://doi.org/10.3758/s13428-014-0527-2).
- *Van Doorn, J.R., Sims, D., & Van Doorn, J. (2015). Scholarly leaders: Value-based leadership and linkages with emotional intelligence and creativity types. *Quarterly Review of Business Disciplines*, 2(2), 83–101.
- Vicent, M., Rubio-Aparicio, M., Sánchez-Meca, J., & González, C. (2019). A reliability generalization meta-analysis of the child and adolescent perfectionism scale. *Journal of Affective Disorders*, 245, 533–544; doi: [10.1016/j.jad.2018.11.049](https://doi.org/10.1016/j.jad.2018.11.049).
- Viechtbauer, W. (2010). Metafor meta-analysis package for R, Version 1.4-0 [Computer software]. Retrieved from <http://CRAN.R-project.org/packag=metafor>
- *Wang, T. H., Lin, H. C. K., Wu, T. T., & Huang, Y. M. (2020). A Multimethod approach for supporting reflection and creativity in online collaborative courses. *Journal of Internet Technology*, 21, 1097–1106.
- *Yu, H., Zuo, S., Liu, Y., & Niemiec, C.P. (2021). Toward a personality integration perspective on creativity: Between- and within-persons associations among autonomy, vitality, and everyday creativity. *The Journal of Positive Psychology*, 16, 789–801; doi: [10.1080/17439760.2020.1818810](https://doi.org/10.1080/17439760.2020.1818810).
- *Zhao, Y., Shen, W.B., Shi, C.H., & Li, M.C. (2017). *The influence of gender and scientific program experience on postgraduates' creativity* [Conference session]. Suzhou, China: MSHH 2017 International Conference on Medical Science and Human Health.
- Zhou, J., & George, J.M. (2001). When job dissatisfaction leads to creativity: Encouraging the expression of voice. *Academy of Management Journal*, 44, 682–696; doi: [10.2307/3069410](https://doi.org/10.2307/3069410).

Sedat Sen, Harran University

Süreyya Yörük, Marmara University

Correspondence concerning this article should be addressed to

Sedat Sen, Department of Educational Sciences, Harran University, Sanliurfa, Türkiye. E-mail: sedatsen@harran.edu.tr

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. Cronbach's α coefficients included in the analyses and characteristics of the studies.

Table S2. A Summary of the categorical study characteristics.

Table S3. A Summary of the continuous study characteristics and α values

Table S4. The differences between the inducing and reporting studies.

Figure S1. Funnel plot of the Bonett's transformed alpha coefficient for the K-DOCS Total score.

Figure S2. Funnel plot of the Bonett's transformed alpha coefficient for the self/everyday subscale.

Figure S3. Funnel plot of the Bonett's transformed alpha coefficient for the scholarly subscale.

Figure S4. Funnel plot of the Bonett's transformed alpha coefficient for the performance subscale.

Figure S5. Funnel plot of the Bonett's transformed alpha coefficient for the mechanical/scientific subscale.

Figure S6. Funnel plot of the Bonett's transformed alpha coefficient for the arts subscale.

Figure S7. Forest plot displaying the alpha coefficients (and 95% confidence intervals) for the K-DOCS Total score.

Figure S8. Forest plot displaying the alpha coefficients (and 95% confidence intervals) for the Self/Everyday subscale.

Figure S9. Forest plot displaying the alpha coefficients (and 95% confidence intervals) for the Scholarly subscale.

Figure S10. Forest plot displaying the alpha coefficients (and 95% confidence intervals) for the Performance subscale.

Figure S11. Forest plot displaying the alpha coefficients (and 95% confidence intervals) for the Mechanical/Scientific subscale.

Figure S12. Forest plot displaying the alpha coefficients (and 95% confidence intervals) for the Artistic subscale.